# Post-pandemic Recovery: investigating Factors that Affected Students' Online Engagement during the Pandemic in Ghana – A Machine Learning Approach

Delali Kwasi Dake
Department of ICT Education
University of Education, Winneba

## ABSTRACT
The COVID-19 pandemic remains a scar on humans, with profound lessons for the future. The post-pandemic recovery in education is crucial for guiding proactive measures for educational stakeholders and governmental entities. During the pandemic, in-person teaching and learning was halted with hasty migration to online learning, especially in developing countries. Students in developing and underdeveloped nations riskily adapted to complete the academic semester online amid concerns of getting the infectious virus. Instructors' pedagogy abruptly shifted, with little or no training on professional approaches to support students during online learning in the face of a virus. This research employed a machine learning approach to examine online education during the pandemic in anticipation of unforeseen circumstances. K-modes clustering was applied to a categorical dataset to reveal latent cluster formations and variable correlations. The naïve bayes classifier was then used to build a predictive model for future cluster members. Four clusters were established, encompassing patterns of unsupportive parents, inadequate internet connectivity, supportive instructors, and average academic success. The naïve bayes classifier also outperformed the random forest algorithm with an accuracy of 84.47%.

## Keywords
K-modes, COVID-19 pandemic, classification, educational data mining, education 4.0, correlation heatmap in education.

## 1. INTRODUCTION
The recent Covid-19 epidemic and Ebola outbreak impacted the educational landscape globally with huge insights on changes and policies for the future [1]–[3]. The virus's contagious nature and rapid spread forced the closure of most educational institutions, as well as lockdowns in many countries. Even affluent nations abandoned in-person teaching, with the primary goal of preserving lives [4], [5]. The repercussions of the economic collapse and difficulties during the pandemic severely impacted developing nations, plunging the educational sector into a crisis [6], [7]. The impulsive response by educational stakeholders encountered infrastructural, network, technical, and pedagogical obstacles during the rapid shift to online learning [8]. The pandemic threatened the scheduled graduation of students, with several institutions finding it difficult to complete the semester [8], [9] The abrupt closing of schools resulted in educational disparities with significant consequences for equity, fairness, and social justice [5], [10]. Students residing in remote places with inadequate internet connectivity, unsupported devices, and unfavourable learning environments face even greater challenges [11], [12]. The pandemic revealed vulnerabilities in the educational sector while simultaneously presenting an opportunity for the implementation of proactive policies moving forward.

Ghana was not immune to the worldwide problems posed by the pandemic. On March 16, 2020, the President of the Republic of Ghana declared the closure of all basic schools, senior high schools, and universities across the nation. The President instructed the ministry of education and communication to establish a support mechanism for the continuation of teaching and learning online [48], [49]. The President's directive was met with strict compliance, which later created financial turbulence for private institutions because most students in the private sector defaulted on fee payments. An estimated 126 private schools collapsed during the pandemic due to salary and Social Security and National Insurance Trust (SSNIT) payment difficulties [50], [51]. Students in public and private institutions, from elementary to tertiary, after the President's announcement must find safe ways to return home with pending psychological and mental health challenges while preparing urgently to switch to online learning [13]–[15]. The stay-at-home and study restrictions imposed during the pandemic necessitated a support system for the students from their teachers, siblings, spouses, and parents. Teachers needed to modify pedagogy to incorporate empathy, counselling, and emotional intelligence into the teaching process. Parents, siblings, and spouses needed to establish a friendly environment at home with minimal interruption during online engagement [16]–[18]

Machine learning (ML) has proved effective in identifying hidden patterns in data through the use of algorithms. ML has three primary categories: supervised learning, unsupervised learning, and reinforcement learning. As shown in Figure 1, supervised learning trains data based on input and output labels, unsupervised learning finds clusters in data without class labels, and reinforcement learning relies on trial-and-error functionality to build rewards using agent actions [19]–[21].

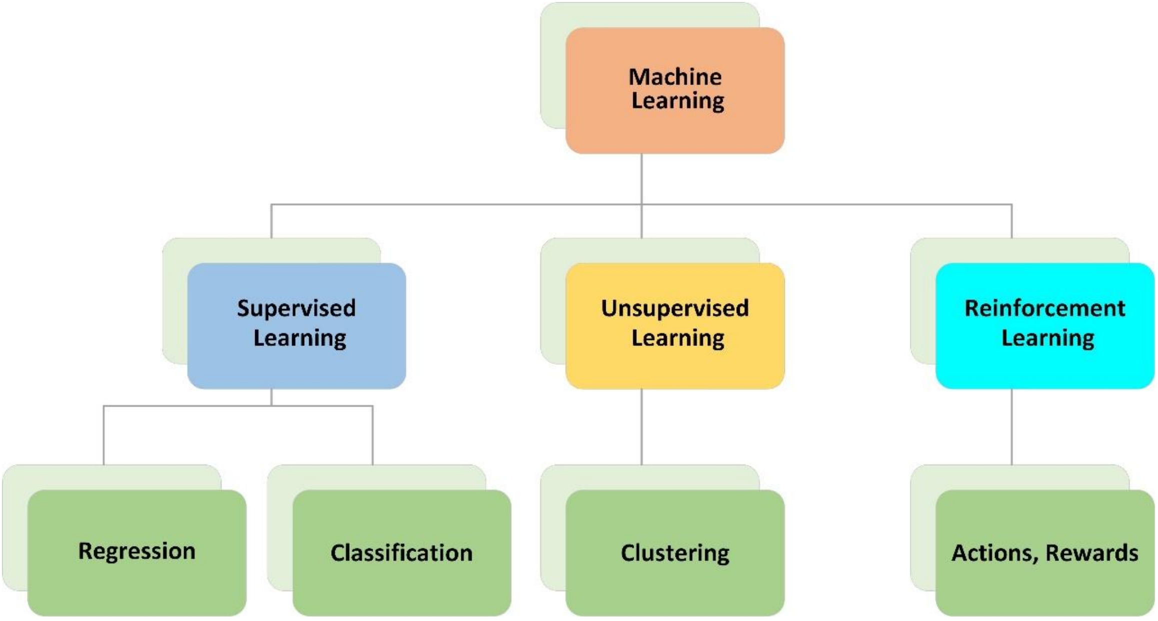


**Fig 1: Types of Machine Learning Algorithm [24]**

Several sectors have used machine learning to automate processes and uncover relationships between variables for policy integration and growth [22], [23]. In healthcare, ML has been used in personalised medicine, clinical trials and research, epidemic forecasting, medication manufacture and discovery, patient behaviour monitoring, robotic-assisted surgery, medical imaging, and hospital equipment management [24], [25]. In industry, ML has seen significant usage in predictive maintenance, increased productivity and efficiency, intelligent decision-making, fuelling robots technology, routine automation, cost reduction, and enhanced product quality [26], [27]. In agriculture, ML has been applied in crop monitoring, farm diseases detection, price prediction with market forecasting, yield optimisation, soil texture and nutrient deficiency detection, and automated irrigation [28], [29]. In banking and finance, ML application is intense in money laundering detection, risk detection and management, automated customer loan system, operation efficiency, personalised banking, and customer chatbots [30]–[32]. As illustrated in Figure 2, the educational sector has seen enormous application paradigm of machine learning as well.

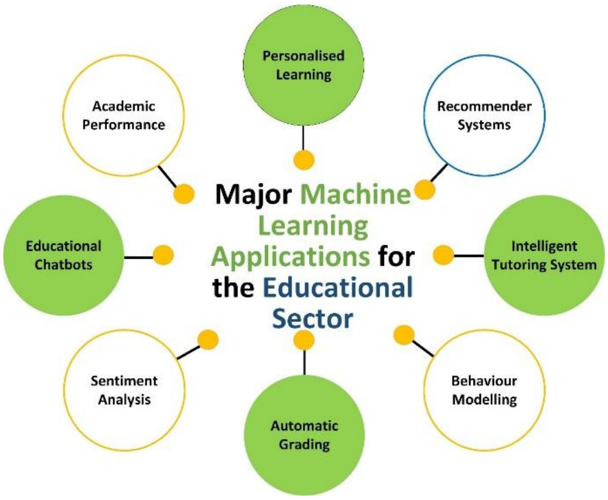


**Fig 2: Machine learning in Education**

ML usage in education has been extremely relevant in personalised learning, academic performance prediction, educational chatbots, sentiment analysis, automatic grading, behaviour modelling, intelligent tutoring system, and recommender systems [33]–[35].

The fundamental goal of this research is to apply machine learning approaches to uncover student patterns during the pandemic and define variable correlations that influenced learners' online engagement. The study will aid educational stakeholders in Ghana in formulating policy measures for unforeseeable future events that may impact teaching and learning. In line with the objectives, we pose the following research questions (RQs) to guide the study:

(1.) How many cluster groupings were formed during online learning when in-person teaching was discontinued because of the pandemic

(2.) What are the similarities and differences between each cluster groupings

(3.) What is the correlation among the features of the dataset

(4.) Which algorithm performed best in building a classification model for the clusters formed.

The rest of the study has this structure: Chapter 2 addresses reviews regarding factors that affected students during the pandemic. Chapter 3 dives into the methodology used in the study. Chapter 4 covers simulation and outcomes. Chapter 5 discusses and analyses the results. Chapter 6 details the conclusions, limitations, and future research.

## 2. LITERATURE REVIEW

The literature review focusses on the experiences of students and teachers during the pandemic-induced school closures in Ghana, which required the transition to online learning from home. The review comprises two primary sections: the basic research component and machine learning. The basic review investigates students' and teachers' experiences with online learning during the pandemic, as well as the application of statistical methods for analysis. The machine learning component comprises the use of algorithms to rank and identify factors that influenced online education and learning during the pandemic.

## 2.1 Basic studies

The papers examined pertain to pedagogy using sample sizes during the pandemic. The primary objective of the basic research is to establish the experiences of students and teachers throughout the pandemic and evaluate the efficacy of online learning while confined at home. Sarpong et al. [36] examined the perceptions of 2000 students from Kumasi Technical University in Ghana on e-learning during the pandemic. The findings reveal that around 90.1% of the students were dissatisfied with the use of e-learning during the pandemic. The students identified unpredictable electricity, unreliable internet connectivity, inadequate infrastructure support, expensive internet data costs, and limited access to technology gadgets as significant obstacles. Agormedah et al. (2020) collected data from 467 students at the University of Cape Coast, Ghana, concerning their online learning experiences during the pandemic. Even though 76.4% of students are familiar with online learning and the requisite gadgets, the majority reported a lack of training in utilising online learning platforms. The students expressed significant dissatisfaction with internet availability and the exorbitant cost of internet bundles as major obstacles to online learning during the pandemic. Owusu-Fordjour et al. [52] investigated the impact of COVID-19 on online learning in Ghana. The data includes 214 respondents from senior high school, college, and university. A significant proportion of the students (88.9%) indicated that their parents were unable to assist them in using the internet for online learning. Regarding parental support, 82.2% reported that their parents were unsupportive at home. Furthermore, 77.6% of the respondents expressed frustration regarding the insufficiency of appropriate learning resources to facilitate self-learning. Overall, 81.3% of respondents disagreed with the claim that they were able to learn effectively at home during the pandemic. Addae et al. [37] examined the digital learning experience of 15 conveniently samples students from a college in Ghana. Their research was mainly qualitative, with learners able to articulate their experiences during the pandemic. The enquiries pertained to the digital learning tools utilised during the lockdown, student perspectives on improving their digital learning experiences, and the difficulties encountered when using the digital platforms. The findings reveal that instructors' empathy, elevated data cost, unreliable internet, and the utilisation of unregulated social media sites significantly impacted their digital learning experience. Kumi et al. [38] investigated 218 allied health sciences students in their second, third, and fourth years at the University of Ghana regarding their remote learning experiences during the pandemic. The results indicate that 93.6% of the students have computing devices for the online learning. Furthermore, 53.1% of the students perceived themselves as moderately independent learners with significant responsibility. Even though the majority of the learners were pleased with remote learning, there was a moderate presence of general anxiety disorder among the students during the pandemic. Ofori Atakorah et al. [39] collected data on 198 students from the seventh-day adventist college of education in agona-ashanti regarding the online learning challenges they experienced during the pandemic in Ghana. Most of the students according to their findings were using android phones during online lectures. In addition, WhatsApp, Zoom Cloud Meeting, and Google Classroom were the predominant platforms utilised for online learning. Their results show that the pandemic has a negative effect on the students with more than 70% of the learners complaining about internet connectivity, insufficient data, and high cost of internet services. Figure 3 depicts the summary of literature review on basic studies regarding students perception about the pandemic effect on online learning in Ghana.

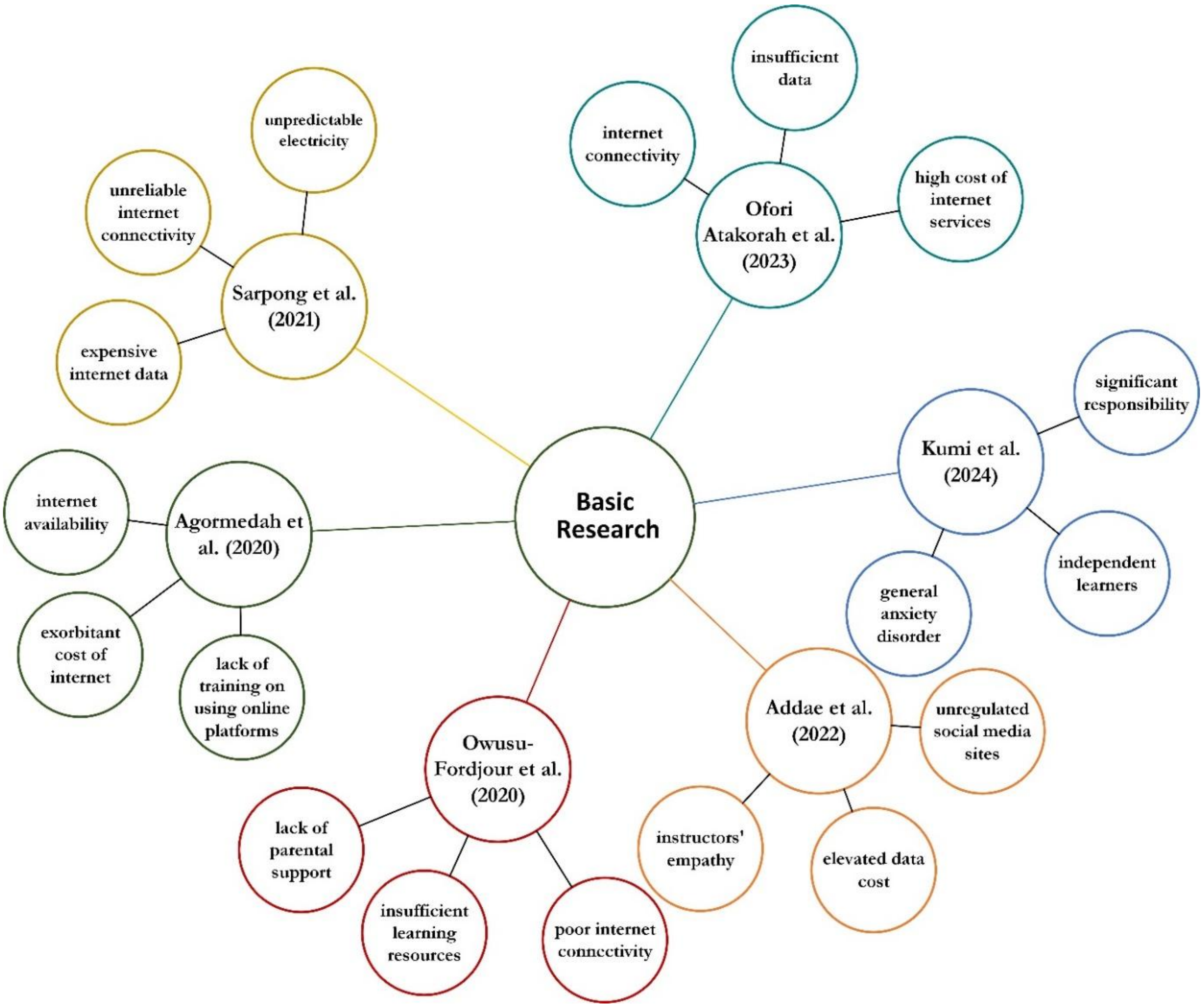


**Fig 3: Summary of Literature Review on Basic Studies**

## 2.2 Machine learning studies

The machine learning research on online education during the pandemic predominantly examined learners' sentiments and the adoption of mobile learning platforms throughout the crisis. There was no literature in Ghana on factor determination and cluster analysis during the pandemic when students were learning online. Almaiah et al. [40] employed the machine learning technique to analyse variable correlation among student acceptance of mobile learning applications during the pandemic. The Random Forest (RF), and the IBK algorithms performed with the highest accuracy of 87.06%. The study also found that the constructs perceived ease of use (PEOU), perceived usefulness (PU), perceived enjoyment (PE), efficiency, effectiveness, and behavioural intention all predict the acceptance of mobile learning. Mujahid et al. [41] investigated 17,155 tweets regarding online education during the pandemic. Their study analyses the sentiments of students, teachers, and stakeholders to ascertain the effectiveness of online lessons during the pandemic. The researchers evaluated the effectiveness of traditional machine learning algorithms against deep learning techniques utilising Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction methods. The Random Forest (RF) and the Support Vector Machine (SVM) achieved the highest accuracy of 95% using BoW. Khattar et al. [42] used machine learning algorithms to investigate learning style, e-learning, and mental health of 583 Indian students. A majority of students (69.8%) expressed dissatisfaction with the lockout due to the inability to interact with their peers in-person, significantly impacting their social lives and online engagement. Students were psychologically unprepared for online learning due to inconsistent sleep patterns, fear of infection, reduced physical activity, and lack of personal space. They used association rule in machine learning to determine the occurrence of these psychological factors during the online learning. The results indicate that 20 rules out of 1552 show high level of confidence and are reliable. The association rule indicates that factors such as the absence of peers, disorganised online learning, continuous computer usage, and extended screen time influenced their engagement during online course delivery. Asad et al. [43] implemented machine learning-based hybrid ensemble methods to determine features that will help students perform better online during the pandemic. Their hybrid ensemble model, which includes decision tree (DT), naïve Bayes (NB), k-nearest neighbour (KNN), support vector machine (SVM), and linear regression (LR), attained an accuracy of 98.6%. The predictions indicated that 12.91% of online learners were at risk, while 87.09% were deemed safe. Altaf et al. [44] used a hybrid framework of deep learning techniques to predict the online performance of Pakistan students during the pandemic. The findings indicate that the proposed hybrid of ID convolutional neural networks (CNNs) and long short-term memory (LSTM) attained an accuracy of 98.8%. Their findings also suggest that some predominant factors influencing learners' online performance include quizzes, assignments, class engagement, internet access, mental wellness, and sleep duration.

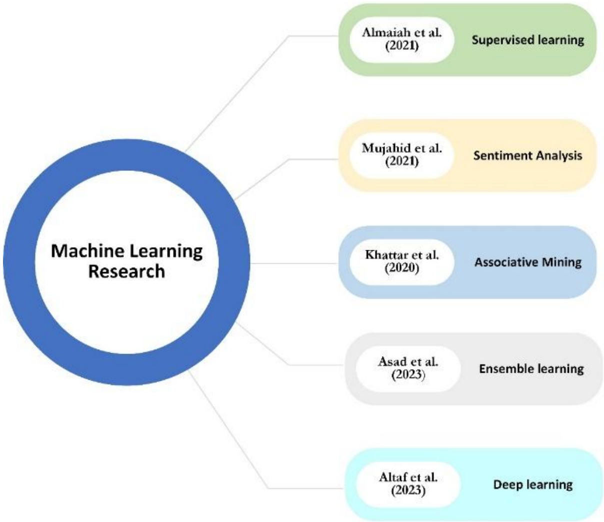


**Fig 4: Summary of Literature Review on Machine Learning Studies**

## 3. RESEARCH METHODOLOGY

The study methodology depicted in Figure 5 includes data collection, data pre-processing, unsupervised learning, supervised learning, feature engineering, and results evaluation.

### 3.1 Student data

The data was collected from level 200 and 300 students in the Department of ICT Education, University of Education, Winneba in 2021 during the pandemic. The convenient sampling, a non-probability sampling technique employed mostly due to the convenience of access to respondents online, was chosen. The total number of valid responses that were recovered for analysis was 536, out of a total population of 850. A Google form was used to administer the questionnaire online, and rigorous confidentiality measures were taken to ensure that the respondents' information remained private with non-disclosure agreement.

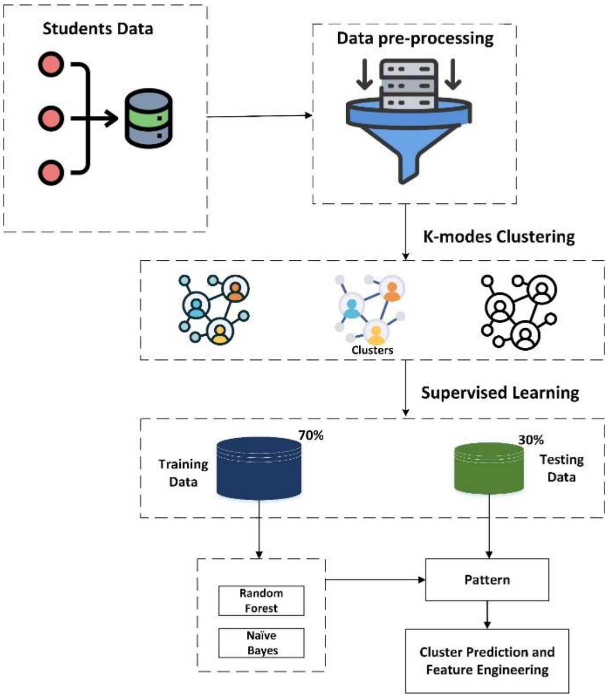


**Fig 5: Methodological Flow of Study**

Even though the responses could be traced back to the students because of the extract of their index number and end-of-semester exam scores, students were requested to agree to a consent form before filling the questionnaire

### 3.2 Data description

Table 1 delineates the demographics of the respondents. Concerning gender, 89.93% of the participants were males, whilst 10.07% were females. The age distribution includes young learners under 22 years (19.03%), individuals aged 23 to 27 (40.30%), and mature students over 27 years (40.67%). The profile indicates that 22.20%, representing the largest proportion of responders, were located in the central area of Ghana during the pandemic. The least represented regions are the Savannah and North East, each with a representation percentage of 0.93

**Table 1. Demographic profile of students**

| Questions | Options | Respondents | Percentage |
|---|---|---|---|
| Gender | Male | 482 | 89.93 |
| | Female | 54 | 10.07 |
| Age | Young (Below 22) | 102 | 19.03 |
| | Moderate (23 – 27) | 216 | 40.30 |
| | Matured (Above 27) | 218 | 40.67 |
| Region of abode during the COVID-19 pandemic | Ashanti | 71 | 13.25 |
| | Bono | 32 | 5.97 |

| | | | |
|---|---|---|---|
| | Bono East | 27 | 5.04 |
| | Ahafo | 8 | 1.49 |
| | Central | 119 | 22.20 |
| | Eastern | 51 | 9.51 |
| | Greater Accra | 67 | 12.50 |
| | Northern | 17 | 3.17 |
| | Savannah | 5 | 0.93 |
| | North East | 5 | 0.93 |
| | Upper East | 21 | 3.92 |
| | Upper West | 8 | 1.49 |
| | Volta | 49 | 9.14 |
| | Western | 27 | 5.04 |
| | Oti | 16 | 2.99 |
| | Western North | 13 | 2.43 |

Table 2 contains enquiries regarding the respondents' place of abode during the pandemic. It was essential to evaluate students' preparedness for online learning. The enquiries directed at the respondents examines diverse situations, including internet accessibility, computing devices, family collaboration and support, as well as emotional triggers of the pandemic. The findings indicate that the respondents experienced poor internet connectivity and unfavourable learning conditions at home. Nonetheless, the encouragement from siblings and overall family support was significant. Students were also frustrated with the pandemic with 84.89% living in a state of fear of contracting the virus.

**Table 2. Place of abode related questions during online learning**

| Questions | Options | Respondents | Percentage |
|---|---|---|---|
| Internet Connectivity (Home) | Poor | 130 | 24.25 |
| | Satisfactory | 200 | 37.31 |
| | Good | 175 | 32.65 |
| | Excellent | 31 | 5.79 |
| Learning Condition (Home) | Not Conducive | 108 | 20.15 |
| | Satisfactory | 216 | 40.30 |
| | Good | 194 | 36.19 |
| | Excellent | 18 | 3.36 |
| Personal Computer | Yes | 490 | 91.42 |
| | No | 46 | 8.58 |
| | Yes | 385 | 71.83 |

| | | | |
|---|---|---|---|
| Home Facilities for academic work and comfort | No | 151 | 28.17 |
| Siblings Attitude (Do they disturb?) | No | 206 | 38.43 |
| | Medium | 244 | 45.52 |
| | High | 86 | 16.05 |
| Covid-19 and state of mind (Fear, Scared?) | Scared (Bothers me always, always in panic mode) | 190 | 35.45 |
| | Average (Sometimes bothers me, occasionally my mind goes there) | 265 | 49.44 |
| | Good (I am only focused on learning, not bothered at all about contracting the virus) | 81 | 15.11 |
| General Family Support (Parental Encouragement) | Poor | 49 | 9.14 |
| | Average | 313 | 58.40 |
| | Excellent | 174 | 32.46 |

Table 3 presents a summary of the support provided by academics during the pandemic, alongside the overall performance of students in their end-of-semester examinations. In general, instructors provided students with emotional support and extended assistance during the online learning (48.51%). Even though majority of students did not achieve outstanding grades, their performance was nonetheless satisfactory, with a significant number attaining very good and average results.

**Table 3. Lecturers support and academic performance**

| **Questions** | **Options** | **Respondents** | **Percentage** |
|---|---|---|---|
| Lecturers Support (Encouragement and online learning support during the pandemic) | Poor (Bad, no lecturer seems to care) | 19 | 3.54 |
| | Average (Unsatisfactory, only a couple of lecturers support) | 257 | 47.95 |
| | Excellent (Most lecturers support) | 260 | 48.51 |
| Academic Performance of Student during the pandemic | Excellent (80 – 100) | 21 | 3.92 |
| | Very Good (70 – 79) | 206 | 38.43 |
| | Average (50 – 69) | 302 | 56.34 |
| | Fail (0 – 49) | 7 | 1.31 |

## 3.3 K-modes Clustering

K-modes clustering is an unsupervised learning approach that employs a partitioning method similar to K-means. However, the distinction lies in the fact that K-means is most effective with numeric data variables, whereas K-modes is employed in clustering to analyse mainly categorical data [45], [46] The K-modes algorithm employs a dissimilarity metric known as the Hamming distance in lieu of the Euclidean distance [47]. Given that x and y are two categorical data objects defined by m features. The dissimilarity metric $d(x, y)$ is mathematically represented in equation 1.

$$d(x, y) = \sum_{j=1}^{m} \delta(x_j, y_j) \quad (1)$$

where

$$\delta(x_j, y_j) = \begin{cases} 0 & \text{if } x_j = y_j \\ 1 & \text{if } x_j \neq y_j \end{cases}$$

**Algorithm 1** K-modes algorithm [47]

**Input:** data set $U = [x_1, x_2, \cdots, x_n]$; the number of clusters k

**Output:** the clustering results $C = [C_1, C_2, \cdots, C_k]$.

1: Randomly choose k modes;

2: Compute the distances between samples and modes

3: Each sample is partitioned into the nearest cluster according to Eq. (1) and update the k modes;

4: Repeat the above steps until the clustering results is convergent

# 4. EXPERIMENTAL RESULTS

The experiments are performed using Anaconda Python distribution software from Continuum Analytics (Conda) with sklearn, matplotlib pandas, SciPy, and NumPy. The presentation layer of the experiment employed Jupyter Notebook, which integrates markdown text with Python source code to form a canvas.

## 4.1 Cluster formation using elbow method in K-modes

**Research Question 1**: How many cluster groupings were formed during online learning when in-person teaching was discontinued because of the pandemic

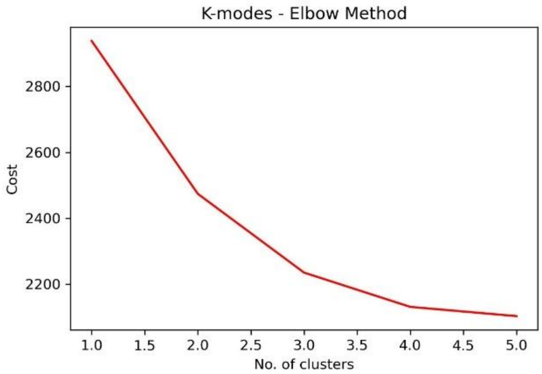


**Fig 6: Cluster formation using Elbow method**

In response to RQ1, the elbow method which graphically determines the ideal number of clusters in K-means, K-modes, and K-prototype algorithms was employed. The elbow is a point of inflection which occurs when the total cluster variance decreases rapidly. As depicted in Figure 6, the point of inflection is selected to be 4. In the diagram above, the number of clusters can either be 3 or 4. This study aims to meticulously evaluate the possible online learning groups that emerged during the pandemic, utilising the maximum point of inflection at 4.

## 4.2 Cluster uniqueness and variations

**Research Question 2**: What are the similarities and differences between each cluster groupings

**Table 4. Characteristics inherent to each cluster**

| Cluster label | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Count | 106 | 66 | 211 | 153 |
| Most common gender | Male | Male | Male | Male |
| Most common age | Moderate (23 – 27) | Matured (Above 27) | Matured (Above 27) | Moderate (23 – 27) |
| Most common region (During Covid-19) | Central | Greater Accra | Central | Ashanti |
| Most common internet connectivity issue (Home) | Poor | Poor | Satisfactory | Good |
| Most common learning condition (Home) | Not conducive | Not conducive | Satisfactory | Good |
| Do most members own personal computer? | Yes | Yes | Yes | Yes |

| | | | | |
|---|---|---|---|---|
| Most common attribute of lecturers during the pandemic | Excellent | Average | Average | Excellent |
| Most common attribute of siblings during the pandemic | Medium | High | Medium | No |
| Do students have home facilities during the pandemic | No | No | Yes | Yes |
| Most common state of mind during the pandemic | Scared | Scared | Average | Average |
| Most common family support during the pandemic | Average | Average | Average | Excellent |
| Most common overall exams performance during the pandemic | Average | Very Good | Average | Average |

In response to RQ2, we used K-modes algorithm to initialise cluster centroids with 100 iterations and a matrix to display cluster members. As shown in Table 4, each cluster has unique cluster variations that is popular among its members. Cluster 0 have the third largest number of members (19.78%), primarily aged between 23 and 27. During the pandemic, they primarily resided in the central region, where they faced poor internet connectivity and an unfavourable learning environment at home. The lecturers provided excellent support to Cluster 0 members, while their siblings and parents provided only average support. Cluster 0 members experienced significant fear during the pandemic, and their exam performance ranged from 50 to 69%. Cluster 1 members have the lowest number (12.31%), and most of them are older than 27. Cluster 1 members during the pandemic mostly stayed in the Greater Accra region, where they had poor internet connectivity and poor learning conditions at home. They faced significant disruptions from their siblings, but were fortunate to have strong parental support. Even though Cluster 1 members lived in fear during the pandemic, they performed very well academically. Cluster 2 have the largest number of members (39.37%) and are mostly above 27 years. Cluster 2 members have stable internet connectivity and a relatively stable online learning condition at home. They received moderate support from their lecturers and parents, while their siblings did not substantially disturb them. They had adequate learning facilities at home and were not overly concerned about the pandemic. The exam performance of Cluster 2 members ranges from 50% to 69%. Cluster 3 has the second largest number of members (28.54%) and resided primarily in the Ashanti region during the pandemic. Cluster 3 members enjoyed stable internet connectivity and good learning conditions at home with little or no interference from their siblings. Cluster 3 members are between ages 23 and 27 with excellent support from their instructors and parents. The pandemic did not emotionally disturb them, and they had all the necessary learning materials at home. However, the performance of cluster 3 members was average (50% to 69%).

## 4.3 Attribute correlation

**Research Question 3**: What is the correlation among the features of the dataset

In response to RQ3, we used the Pearson correlation heatmap to understand the relationship among the categorical attributes. The heatmap, as shown in Figure 7, graphically displays the correlation coefficient between each pair of features. The yellowish colour toward 1 indicates a strong correlation, while the pinkish colour toward -1 indicates a negative correlation. The two-dimensional correlation matrix also indicates that values closer to 0 have no correlation. In general, the heatmap shows no correlation between variables. This largely means that the features are independent of one another. The correlation coefficients of variables from the graph are closer to 0.

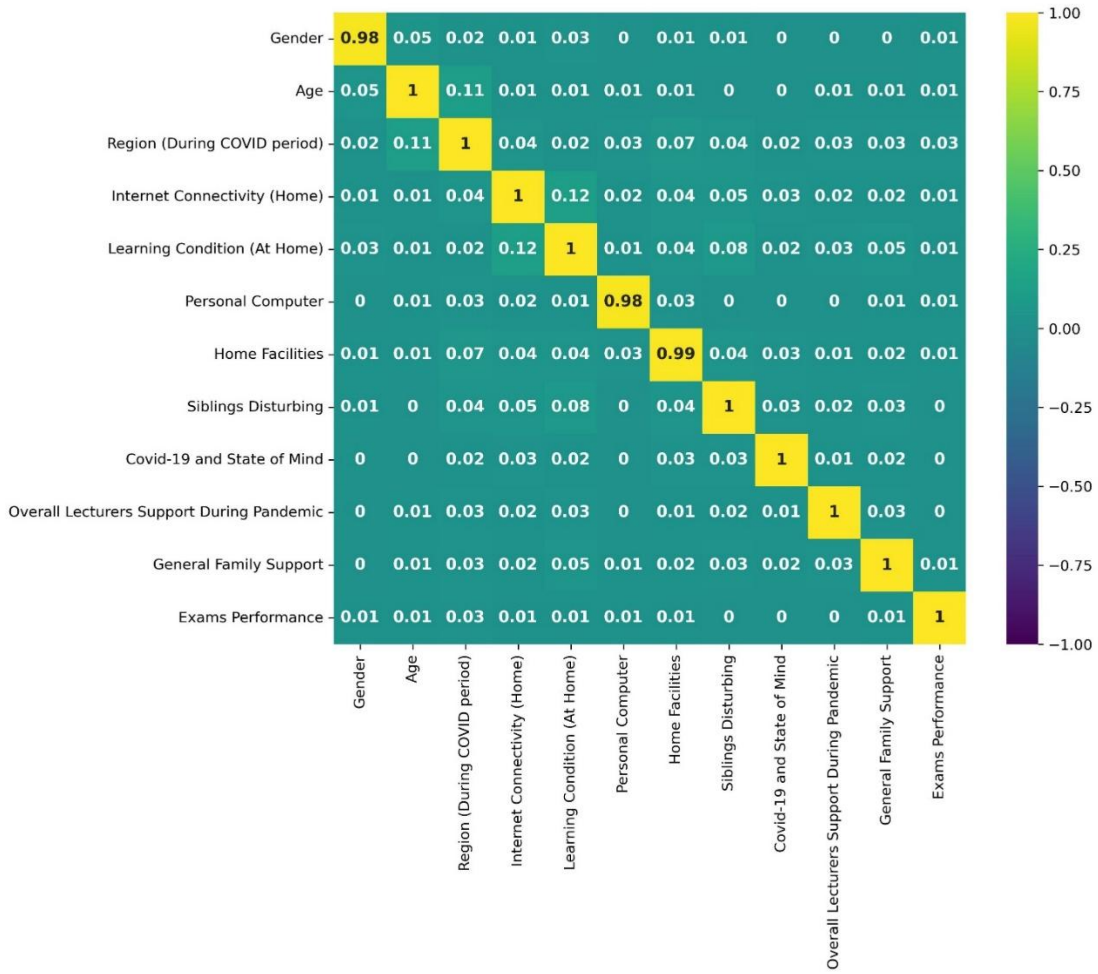


**Fig 7: Pearson correlation heatmap**

## 4.4 Classifiers performance

**Research Question 4**: Which algorithm performed best in building a classification model for the clusters formed.

In addressing RQ4, we assessed the predictive efficacy of the random forest (RF) and naïve Bayes (NB) supervised learning algorithms utilising a 375 (70% training) to 161 (30% testing) split ratio. The subsequent classification metrics were assessed to identify the most effective classifier: accuracy, f-measure, ROC-AUC, and the confusion matrix. Furthermore, we employed the information gain feature engineering method to identify the most influential attributes that substantially impacted the cluster predictions. The findings illustrated in Figure 8 based on the test results indicate that the Naïve Bayes method has the maximum classification accuracy of 84.47%, an F-measure of 0.841, and a ROC-AUC of 0.963.

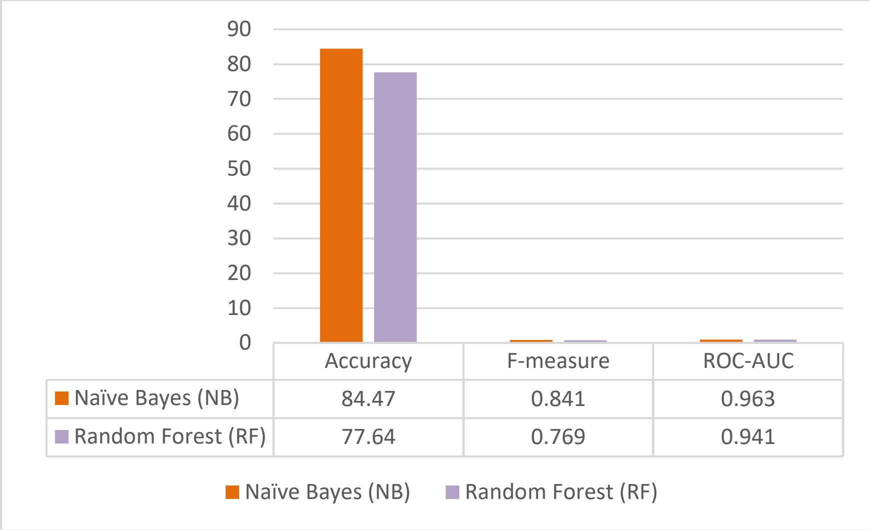


**Fig 8: Classifiers performance**

As illustrated by the confusion matrix in Figure 9, the naïve Bayes classifier correctly predicted 136 test samples, with 25 incorrectly predicted. The random forest, on the other hand, predicted correctly 125 test samples as compared to 36 misclassified samples.


**Fig 9: Confusion matrix of classifiers. (a) Naïve Bayes (b) Random Forest**

As shown in Figure 10, the feature ranking test demonstrates that the learning condition at home, internet connectivity, sibling disturbance, and region of residence during the pandemic all had an impact on building the classification model for the clusters. The least important attributes comprise gender, personal computer usage, examination performance, and age. This indicates that those attributes have minimal or no impact on constructing the classification model

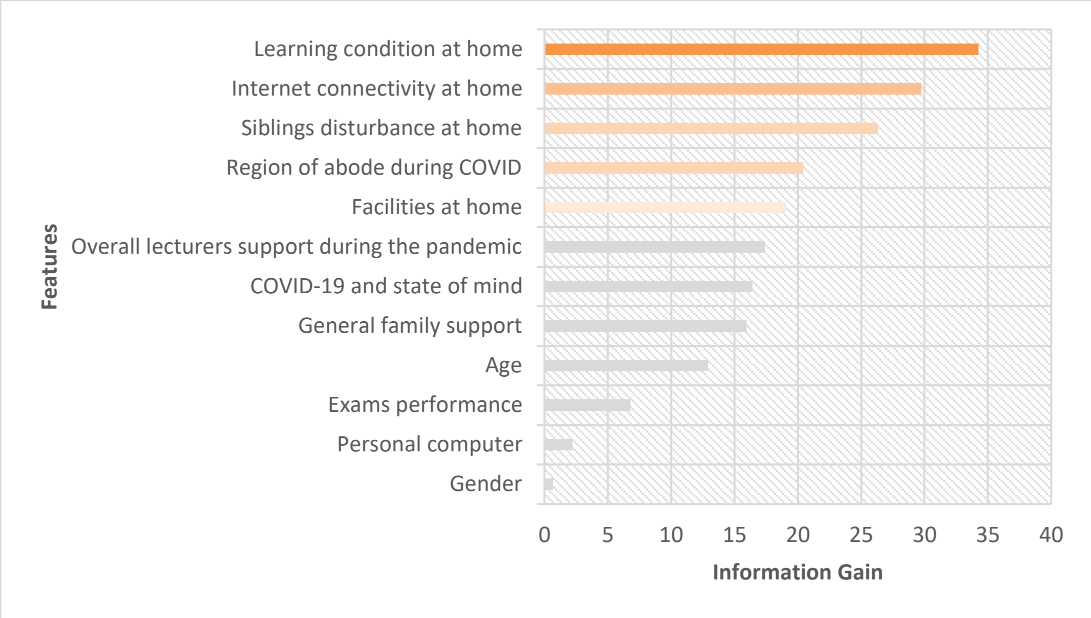


**Fig 10: Feature relevance in descending order**

## 5. FINDINGS AND SUMMARY

This study revealed a hidden pattern in students' online learning during the pandemic, displaying intriguing cluster formation and patterns. During the pandemic, most institutions around the world, including those in Ghana, halted in-person teaching in order to limit the transmission of the infectious virus. Students, while interacting with their instructors at home, encounter scenarios that require analysis to inform institutional and governmental policies for proactive adoption. The research collected data from students in 2021 when traditional classroom teaching and learning was suspended at the University of Education, Winneba, necessitating the completion of the semester online and remotely. Even though the pandemic effect has declined significantly in 2024, there was the need to understand factors that affected learners performance and provide a policy framework. The findings from this study, summarised with the descriptive statistics, primarily on internet connectivity, remote learning support, inconvenient learning environment and learner frustrations, were also found in the reviewed literature. The Owusu-Fordjour et al. (2020) study reveals that parents were largely unsupportive during the pandemic regarding online learning. Addae et al. [37] also referenced a deficiency in teachers' empathy, inadequate internet access, and elevated data prices as problems faced by students during online learning throughout the pandemic. Ofori Atakorah et al. [39] study also emphasised internet connectivity, insufficient data, and high cost of internet services as primary challenges faced by learners during online learning in Ghana.

The second major aspect of this study used the K-modes algorithm to cluster students during the pandemic and reveal the uniqueness of each cluster for national and proactive policies. The students' residential region, age, familial support, internet access, and examination performance indicate a discernible pattern. Cluster 3 members have a conducive environment during the pandemic but performed averagely in the exams. Generally, this cluster group will need additional academic support when in-person teaching and learning resumes. Cluster 1 members as the least represented, consist mostly of matured students. Even though they were disadvantaged by having to take care of their siblings while continuously living in fear because of the pandemic, their academic performance was good. Cluster 1 members can spearhead projects and team initiatives upon the resumption of in-person instruction. They can serve as mentors for students with academic difficulties and offer support to the department's counselling unit. Comparatively, there is no study in literature that utilised K-modes for categorical data clustering in education during online learning in Ghana. Aside clustering, the study built a classifier to predict the cluster groupings of students. The Naïve Bayes (NB) classifier outperformed the Random Forest (RF) algorithm for all the major machine learning metrics. The Naive Bayes algorithm attained a test accuracy of 84.47%, compared to 77.64% for the Random Forest. The classification results, however, differ significantly from research in literature that utilised machine learning for students' online engagement during the pandemic. While previous studies on online learning have used machine algorithms to predict the online performance of students and at-risk learners [43],[44], this study predicted cluster groupings of new learners for proactive academic groupings and tailored support in-case of unforeseen outbreaks.

The study also used the information gain feature engineering mechanism to determine the most influential features for cluster groupings. The results reveal that learning conditions at home, internet connectivity, and family support were crucial features in developing the classification model. This falls in line with Altaf et al. [44] hybrid model with feature engineering during online learning. Their findings show that assignments, class participation, and internet access are crucial factors to consider when predicting students' online performance.

## 6. CONCLUSIONS, LIMITATIONS, AND FUTURE STUDIES

This study reveals intriguing patterns that occurred during the pandemic, when in-person teaching was discontinue at the University of Education, Winneba. This is the first time in Ghana that the K-modes clustering using categorical dataset uncovered cluster patterns of learners' academic difficulties since online learning became required. The cluster groupings will help in personalised and group learning methods as teachers' pedagogy evolves in the future. The cluster groupings revealed struggling online learners even with maximum support and internet connectivity. Genuine concerns regarding disadvantaged students facing adverse remote learning conditions and inadequate support present an opportunity for counselling and increased efforts by educators to avert students' failure. Government investment in internet accessibility and infrastructure initiatives should be equal and fairly distributed throughout the 16 regions of Ghana.

Even though, the study uncovered patterns, an improved data size and residential address location of students during the pandemic will help educational policy makers and government in generalisation and conclusions. The results of this study, derived from the sample size and demographic data, do not fully represent conditions that may exist at other universities in Ghana. Future research will replicate different variables and evaluate students' frustration during the pandemic using sentiment analysis from other prominent universities in Ghana.

## 7. REFERENCES

[1] S. Yunus, M. Mariyudi, and M. Bin Abubakar, "Strategies for optimizing learning activities during the pandemic and new normal," *Cogent Soc. Sci.*, vol. 9, no. 1, 2023, doi: 10.1080/23311886.2023.2175491.

[2] D. DeMatthews, P. Reyes, J. Solis Rodriguez, and D. Knight, "Principal Perceptions of the Distance Learning Transition During the Pandemic," *Educ. Policy*, vol. 37, no. 3, pp. 653–675, 2023, doi: 10.1177/08959048211049421.

[3] M. Babbar and T. Gupta, "Response of educational institutions to COVID-19 pandemic: An inter-country comparison," *Policy Futur. Educ.*, vol. 20, no. 4, pp. 469–491, 2022, doi: 10.1177/14782103211021937.

[4] I. W. Suryasa, M. Rodríguez-Gámez, and T. Koldoris, "Post-pandemic health and its sustainability," *Int. J. Health Sci. (Qassim).*, vol. 6, no. 1, pp. i–v, 2022, doi: 10.53730/ijhs.v6n1.5949.

[5] A. Tarabini, "The role of schooling in times of global pandemic: a sociological approach," *Int. Stud. Sociol. Educ.*, vol. 31, no. 4, pp. 427–445, 2022, doi: 10.1080/09620214.2021.1966825.

[6] S. K. Thaheem, M. J. Zainol Abidin, Q. Mirza, and H. U. Pathan, "Online teaching benefits and challenges during pandemic COVID-19: a comparative study of Pakistan and Indonesia," *Asian Educ. Dev. Stud.*, vol. 11, no. 2, pp. 311–323, 2022, doi: 10.1108/AEDS-08-2020-0189.

[7] D. Willies, "African Journal of Education and Practice ISSN 2519-0296 ( Online ) THE IMPACT OF COVID-19 PANDEMIC ON THE EDUCATION SYSTEM IN DEVELOPING COUNTRIES David Willies African Journal of Education and Practice ISSN 2519-0296 ( Online )," *African J. Educ. Pract.*, vol. 9, no. 1, pp. 15–26, 2023.

[8] P. M. Hash, "Remote Learning in School Bands During the COVID-19 Shutdown," *J. Res. Music Educ.*, vol. 68, no. 4, pp. 381–397, 2021, doi: 10.1177/0022429420967008.

[9] U. Hoadley, "Schools in the Time of COVID-19: Impacts of the Pandemic on Curriculum," *Resep Non-Economic Work. Pap.*, no. November, p. 29, 2020, [Online]. Available: https://www.researchgate.net/publication/345310078_Schools_in_the_Time_of_COVID-19_Impacts_of_the_Pandemic_on_Curriculum.

[10] J. H. L. Koh and B. K. Daniel, "Shifting online during COVID-19: A systematic review of teaching and learning strategies and their outcomes," *Int. J. Educ. Technol. High. Educ.*, vol. 19, no. 1, 2022, doi: 10.1186/s41239-022-00361-7.

[11] N. E. Mahaye, "The Impact of COVID-19 Pandemic on Education : Navigating Forward the Pedagogy of Blended Learning," *impact covid-19 pandemic Educ. Navig. Forw. Pedagog. Blended Learn.*, no. April, p. 3, 2020, [Online]. Available: https://www.researchgate.net/publication/340899662.

[12] S. Tadesse and W. Muluye, "The Impact of COVID-19 Pandemic on Education System in Developing Countries: A Review," *Open J. Soc. Sci.*, vol. 08, no. 10, pp. 159–170, 2020, doi: 10.4236/jss.2020.810011.

[13] K. Durizzo, E. Asiedu, A. van der Merwe, and I. Günther, "Economic Recovery but Stagnating Mental Health During a Global Pandemic? Evidence from Ghana and South Africa," *Rev. Income Wealth*, vol. 68, no. 2, pp. 563–589, 2022, doi: 10.1111/roiw.12587.

[14] K. Oppong Asante, E. N. B. Quarshie, and J. Andoh-Arthur, "COVID-19 school closure and adolescent mental health in sub-Saharan Africa," *Int. J. Soc. Psychiatry*, vol. 67, no. 7, pp. 958–960, 2021, doi: 10.1177/0020764020973684.

[15] M. Aarah-Bapuah, S. Sarpomaa Oppong, A. Ohenewaa Yawson, G. Dzansi, and S. Adjorlolo, "Covid-19 and mental health of children and adolescents: A systematic review," *Cogent Psychol.*, vol. 9, no. 1, 2022, doi: 10.1080/23311908.2022.2111849.

[16] S. Wolf, E. Aurino, N. M. Suntheimer, and E. A. Avornyo, "Optional ERIC Coversheet — Only for Use with U . S . Department of Education Grantee Submissions This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form is not included on the PDF to," 2022.

[17] E. K. AGORMEDAH, E. ADU HENAKU, D. M. K. AYİTE, and E. APORİ ANSAH, "Online Learning in Higher Education during COVID-19 Pandemic: A case of Ghana," *J. Educ. Technol. Online Learn.*, vol. 3, no. 3, 2020, doi: 10.31681/jetol.726441.

[18] F. Ansah *et al.*, "Households' level considerations in the vulnerability to potential learning loss among school children in Ghana," *Equity Educ. Soc.*, no. July 2023, 2023, doi: 10.1177/27526461231194787.

[19] M. Javaid, A. Haleem, R. Pratap Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications," *Int. J. Intell. Networks*, vol. 3, no. February, pp. 58–73, 2022, doi: 10.1016/j.ijin.2022.05.002.

[20] S. Gu *et al.*, "A Review of Safe Reinforcement Learning: Methods, Theories and Applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 1–20, 2024, doi: 10.1109/tpami.2024.3457538.

[21] K. K. Verma, B. M. Singh, and A. Dixit, "A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system," *Int. J. Inf. Technol.*, vol. 14, no. 1, pp. 397–410, 2022, doi: 10.1007/s41870-019-00364-0.

[22] Z. Fan, Z. Yan, and S. Wen, "Deep Learning and Artificial

Intelligence in Sustainability: A Review of SDGs, Renewable Energy, and Environmental Health," *Sustain.*, vol. 15, no. 18, 2023, doi: 10.3390/su151813493.

[23] A. Raihan, "A comprehensive review of artificial intelligence and machine learning applications in energy consumption and production," *J. Technol. Innov. Energy*, vol. 2, no. 4, pp. 1–26, 2023, doi: 10.56556/jtie.v2i4.608.

[24] Q. An, S. Rahman, J. Zhou, and J. J. Kang, "A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges," *Sensors*, vol. 23, no. 9, 2023, doi: 10.3390/s23094178.

[25] M. Javaid, A. Haleem, R. Pratap Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications," *Int. J. Intell. Networks*, vol. 3, no. May, pp. 58–73, 2022, doi: 10.1016/j.ijin.2022.05.002.

[26] R. Rai, M. K. Tiwari, D. Ivanov, and A. Dolgui, "Machine learning in manufacturing and industry 4.0 applications," *Int. J. Prod. Res.*, vol. 59, no. 16, pp. 4773–4778, 2021, doi: 10.1080/00207543.2021.1956675.

[27] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017, doi: 10.1109/ACCESS.2017.2756872.

[28] S. O. Araújo, R. S. Peres, J. C. Ramalho, F. Lidon, and J. Barata, "Machine Learning Applications in Agriculture: Current Trends, Challenges, and Future Perspectives," *Agronomy*, vol. 13, no. 12, pp. 1–27, 2023, doi: 10.3390/agronomy13122976.

[29] S. Pokhariyal, N. R. Patel, and A. Govind, "Machine Learning-Driven Remote Sensing Applications for Agriculture in India—A Systematic Review," *Agronomy*, vol. 13, no. 9, pp. 1–30, 2023, doi: 10.3390/agronomy13092302.

[30] M. Leo, S. Sharma, and K. Maddulety, "Machine learning in banking risk management: A literature review," *Risks*, vol. 7, no. 1, 2019, doi: 10.3390/risks7010029.

[31] D. Sudhakar Dharrao, A. M. Bongale, S. T. Deokate, D. Doreswamy, and S. Krishna Bhat, "Citation: Sonkavde, Gaurang," 2023, [Online]. Available: https://doi.org/10.3390/ijfs11030094.

[32] D. Boughaci and A. A. K. Alkhawaldeh, "Appropriate machine learning techniques for credit scoring and bankruptcy prediction in banking and finance: A comparative study," *Risk Decis. Anal.*, vol. 8, no. 1–2, pp. 15–24, 2020, doi: 10.3233/RDA-180051.

[33] [33] C. Yang, F. K. Chiang, Q. Cheng, and J. Ji, "Machine Learning-Based Student Modeling Methodology for Intelligent Tutoring Systems," *J. Educ. Comput. Res.*, vol. 59, no. 6, pp. 1015–1035, 2021, doi: 10.1177/0735633120986256.

[34] B. Memarian and T. Doleck, "A scoping review of reinforcement learning in education," *Comput. Educ. Open*, vol. 6, no. March, p. 100175, 2024, doi: 10.1016/j.caeo.2024.100175.

[35] L. Chen, P. Chen, and Z. Lin, "Artificial Intelligence in Education: A Review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020, doi: 10.1109/ACCESS.2020.2988510.

[36] S. A. Sarpong, G. Dwomoh, E. K. Boakye, and I. Ofosua-Adjei, "Online Teaching and Learning Under COVID-19 Pandemic; Perception of University Students in Ghana," *Eur. J. Interact. Multimed. Educ.*, vol. 3, no. 1, p. e02203, 2021, doi: 10.30935/ejimed/11438.

[37] D. Addae, S. Amponsah, and B. J. Gborti, "COVID-19 Pandemic and the Shift to Digital Learning: Experiences of Students in a Community College in Ghana," *Community Coll. J. Res. Pract.*, vol. 46, no. 1–2, pp. 101–112, 2022, doi: 10.1080/10668926.2021.1972364.

[38] R. Kumi, J. Quartey, S. K. Kwakye, and E. D. Oppong, "Students' perceptions and experiences of remote learning amid covid-19 pandemic in Ghana," *Hong Kong Physiother. J.*, vol. 44, no. 2, pp. 103–118, 2024, doi: 10.1142/S1013702524500070.

[39] P. Ofori Atakorah, E. Honlah, P. Atta Poku Jnr, E. Frimpong, and G. Achem, "Challenges to online studies during COVID-19: The perspective of Seventh-day Adventist College of Education students in Ghana," *Cogent Educ.*, vol. 10, no. 1, 2023, doi: 10.1080/2331186X.2022.2162680.

[40] M. A. Almaiah, O. Almomani, A. Al-Khasawneh, and A. Althunibat, *Predicting the Acceptance of Mobile Learning Applications During COVID-19 Using Machine Learning Prediction Algorithms*, vol. 348. Springer International Publishing, 2021.

[41] M. Mujahid *et al.*, "Sentiment analysis and topic modeling on tweets about online education during covid-19," *Appl. Sci.*, vol. 11, no. 18, 2021, doi: 10.3390/app11188438.

[42] A. Khattar, P. R. Jain, and S. M. K. Quadri, "Effects of the Disastrous Pandemic COVID 19 on Learning Styles, Activities and Mental Health of Young Indian Students-A Machine Learning Approach," *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, no. Iciccs, pp. 1190–1195, 2020, doi: 10.1109/ICICCS48265.2020.9120955.

[43] R. Asad, S. Altaf, S. Ahmad, H. Mahmoud, S. Huda, and S. Iqbal, "Machine Learning-Based Hybrid Ensemble Model Achieving Precision Education for Online Education Amid the Lockdown Period of COVID-19 Pandemic in Pakistan," *Sustain.* , vol. 15, no. 6, 2023, doi: 10.3390/su15065431.

[44] S. Altaf, R. Asad, S. Ahmad, I. Ahmed, M. Abdollahian, and M. Zaindin, "A Hybrid Framework of Deep Learning Techniques to Predict Online Performance of Learners during COVID-19 Pandemic," *Sustain.*, vol. 15, no. 15, 2023, doi: 10.3390/su151511731.

[45] S. Wu, D. D. Jiang, and Q. Wang, "HABOS clustering algorithm for categorical data," *Gongcheng Kexue Xuebao/Chinese J. Eng.*, vol. 38, no. 7, pp. 1017–1024, 2016.

[46] R. Madhuri, M. R. Murty, J. V. R. Murthy, P. V. G. D. P. Reddy, and S. C. Satapathy, "Cluster Analysis on Different Data Sets Using K-Modes and K-Prototype Algorithms," *Adv. Intell. Syst. Comput.*, vol. 249 VOLUME, pp. 137–144, 2014, doi: 10.1007/978-3-319-03095-1_15.

[47] S. Xu, S. Liu, J. Zhou, and L. Feng, "Fuzzy rough clustering for categorical data," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 11, pp. 3213–3223, 2019, doi: 10.1007/s13042-019-01012-6.

[48] Citi newsroom (2020) https://citinewsroom.com/2020/03/coronavirus-shut-down-all-schools-until-further-notice-nana-addo-orders/

[49] Ghanaweb (2020) https://www.ghanaweb.com/GhanaHomePage/NewsArchive/Coronavirus-Schools-across-Ghana-close-down-on-Akufo-Addo-s-order-896008

[50] Citi newsroom (2021) https://citinewsroom.com/2021/01/126-private-schools-have-collapsed-since-covid-19-began-gnacops/

[51] Myjoyonline (2021) https://www.myjoyonline.com/covid-19-school-closure-10-months-of-no-fees-no-salaries/.

[52] Owusu-Fordjour, C., Koomson, C.K., & Hanson, D. (2020). "The impact of COVID-19 on learning - the perspective of the Ghanaian Student". *European Journal of Educational Studies*, 7(3). http://dx.doi.org/10.5281/zenodo.3753586