International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I12P228

Volume 70 Issue 12, 296-307, December 2022

Original Article

Data Mining and Visualisation of Basic Educational Resources for Quality Education

Fuseini Inusah[1,2], Yaw Marfo Missah[1], Ussiph Najim[1], Frimpong Twum[1]

[1] *Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.*
[2] *Department of Mathematics and ICT Education, University for Development Studies, Tamale, Ghana.*

[1]*Corresponding Author : finusah@uds.edu.gh*

Received: 25 July 2022 Revised: 13 October 2022 Accepted: 17 December 2022 Published: 24 December 2022

Abstract - *With an increase in educational resources for the growing population, data for Basic Education (BE) is becoming larger, requiring technical data tools to analyze and interpret. This research uses classification and clustering techniques to analyze the data from public schools in Ghana to identify the challenges. Nine (9) data mining algorithms in rapid miner studio 9.10 were used for the analysis to know the most efficient algorithm suitable for the data. These are; Generalized Linear Module (GLM), Naïve Bayes (NB), Logistic Regression (LR), Deep Learning (DL), Decision Tree (DT), Fast Large Margins (FLM), Gradient Boosted Tree (GBT), Random Forest (RF), and Support Vector Machines (SVM). The performance of GBT was seen as more appropriate, and this algorithm's results were presented. Excerpts from the reports are also included in the form of qualitative data. A diagrammatic representation of the interoperability among levels of education for quality education has also been presented. A proposed Neural Network model has been designed for the challenges and solutions. The conclusions draw that addressing the challenges of BE requires educational policy stability and enforcement to maximize resources and minimize the challenges in schools at all levels of education.*

Keywords - *Challenges, Classification, Clustering, Data mining, Educational resources.*

1. Introduction

The core of the education sector in a country is BE. As the foundation for educational growth and development, success at this level is leveraged to achieve success at higher levels of education. This means that the bottlenecks of BE negatively affect the educational progress in the country at all levels. In [1], the exploration of a rural education system concerning the challenges in the availability of resources and resource mapping in China is presented. This is a historic presentation regarding the teaching methodologies the deprived schools are compelled to use in lesson delivery. This affects the quality of education. As a middle-income country, Ghana has struggled to achieve some amount of success in education with significant educational policies [2]. Both the government and Civil Society Organizations (CSO) and the private sectors invest in BE to help in improving quality education [3]. In the paper [4], a decade trend of the challenges and success of education in Ghana is presented regarding the key issues that need attention. This shows the effort put in place by various governments and international bodies in the country to improve BE—public educational institutions which are a government-own attempt to achieve the targets of the education sector.

As a legal responsibility, the Ministry of Education (MoE) conducts an annual census to gather data for the various educational indicators. This data spans access, quality, management and technology in education. Careful identification of the challenges of BE in the country is often challenging. Using graphs and common tools in spreadsheets does not give enough details of the problem. The application of relationships among various levels of education and the interoperability of quality education concerning the availability and use of resources is a challenge. Technical data analytical tools that can specifically identify the challenges of BE in detail are not used. This makes it difficult to accurately identify the challenges of BE in the country, creating both rural and urban schools that lack basic educational resources to enhance effective teaching and learning. In this research, the quality of education regarding the availability of pedagogical tools to enhance effective teaching and learning is considered by using RapidMiner Studio 9.10 as a data mining software to carefully analyze the data to identify the available resources and challenges of education. The data has been grouped into districts, and the institutions are classified into rural and urban schools.

Further grouping into clusters is used to identify well-endowed schools and deprived schools. This is necessary to find the characteristics of each cluster to identify the unique challenges. Generally, the research aims to identify the challenges of BE for a sustainable solution to improve the quality of education. Specifically, the study has two objectives; 1. To accurately determine the obstacles of BE using data mining techniques 2. To identify sustainable solutions for quality education at the basic level through a proposed model.

The formal educational system of Ghana dates back to the colonial era, which is before 1957. When Ghana gained independence in 1957, it had the strongest educational system in Sub-Sahara Africa till the 1970s [5]. The challenges of BE in Ghana have been a thorn in the flesh of stakeholders in the country, and several interventions have been carried out in an attempt to address the challenges [6]. Mainstream schools and special and inclusive educational institutions are deeply affected [7]. The local and international bodies have invested heavily in addressing the challenges of BE to close the gap between rural and urban education [8]. Despite all these interventions carried out, BE is still characterized by numerous challenges making it difficult for the country to see an improvement in the Key Performance Indicators and achieve its targets in the country [9].

2. Related Work

From [9], the Education Strategic Plan 2018-2030 spells out the strategies needed to improve the education sector's performance. Cost and resource allocation to the various levels of education has been carefully done to provide details and help trace the challenges of education. Projections have been made by using the 2016/2017 academic year as a baseline to predict the education sector's performance going forward. Barely over three years after the prediction, the sector's performance is still far below the projected values for each year. This clearly indicates that if measures are not taken, no expected improvement for the indicators will be achieved. The MoE launched the Educational Management Information System (EMIS) to facilitate the collection of educational data for informed decision-making in the country. This was sponsored by the World Bank and had technical support from Harvard University. From the report [10], UNESCO Institute for Statistics (UIS) gives technical support with an application customized for Ghana. This is a conscious effort made by the MoE to build a stronger database to help in planning and decision-making in the education sector. Despite all these efforts, it is still difficult to accurately predict and plan effectively using the data available from the ministry.

From the paper [11], careful analysis of the available data revealed a fluctuation in the population of school-going age. This is an abnormal trend regarding the population growth of a country that does not experience population decay but rather growth. The inaccuracies and inconsistencies affect the results of educational indicators, which are the basis for decision-making. This subsequently affects the efficiency of decision-making as the speed, accuracy and ease of decision-making are affected [12].

2.1. Classification in Datamining

In the paper [13], classification is a machine-learning technique for constructing prediction models from data. This technique is good for getting unknown values/future values of information for decision-making. In classification, each record has attributes that are used in grouping the data according to the class. Classification will therefore help in finding a model for class attributes as a function of the values of other attributes. In doing this, previously unseen records are assigned classes as accurately as possible. The model is developed by training data sets and validated with different test sets of the same data type. According to [14], classification in a dynamic environment is always a challenging task in data mining. The use of training data to get the model and the use of test data to test the accuracy of the model may not be feasible in real-life situations. A comprehensive data mining approach to classification is necessary for dynamic environments. From [15], novelty detection, where abnormal patterns are embedded in larger normal data, is needed to improve the accuracy of dynamic environment classification. Despite these challenges, classification in big data is the most recommended task in predicting future values for decision-making.

2.2. Clustering in Datamining

Unlike classification, in clustering, variables of similar features are put into sub-groups or categories known as clusters. This does not go by pre-defined data guided by the algorithm. Records or rows are clustered based on similarities [16]. The data miner is responsible for assigning meaning to the clusters. In one cluster, the variables are alike and different from that of other clusters. Labels can be tagged to clusters and assigned records in the database to their respective classes after grouping the variables, which can be used for data classifications [32]. The paper's author [18] sees clustering as an independent data mining method used to organize data sets and analyze them using other data mining techniques. Common forms of clusters include; contiguous clusters, density-based clusters, centre-based clusters, Well-separated clusters, and Conceptual clusters or shared properties. In this study, the clustering of schools will be done based on the well-endowed and deprived schools. Further sub-groups or clusters, such as deprived and well-endowed schools, will also be identified within the various classes. It will help in finding the needs of the learners and the teachers in all schools.

2.3. Application of Data Mining in Education

In the paper [19], an intelligent prediction model for student performance is done using an ensemble and filtering approach. This is seen to be more efficient than individual classifiers in generating predictions. Speed in prediction, accuracy and ease to use of the system is seen in predicting the students' performance. As indicated in [20], the core objective of every educational institution is to maximize performance and reduce the level of failure among students.

This can be achieved if there is a proper prediction of the learner's abilities to know the strengths and weaknesses. This will be more effective when early prediction is made using a web-based support system. As indicated in [20], the teacher and the learner's pattern and logs can be monitored to improve learners' academic performance in university through data mining techniques. Clustering, classification and visualization of the data can help to identify the patterns for early detection and correction. In a related paper [21], event logs downloaded from the higher levels of learning in Croatia were analyzed using clustering and decision tree techniques. This was done regarding the decision-making on how students learn with due consideration of the learner's behaviour. A survey on educational data mining by [32] reveals relevant information on modifying the teaching and learning environment to improve effective teaching and learning using Educational Data Mining tools. In [32], a hybrid data mining model is proposed to enhance students' performance and evaluation accuracy. Naïve Bayes and J48 classifier techniques are used in the proposed model to enhance efficiency for accurate analysis of learners' performance in a precise manner. The application of educational data mining to a massive course can be seen in [23]. In a learning management system (LMS) delivered by a Brazilian university, the use of EDM algorithms revealed relevant information, including the attributes of learners as the most significant factor contributing to the learner's performance. Still, in the same country, the academic performance of students in public schools is predicted using data mining techniques. Gradient boosted tree was used in the prediction to enhance efficiency in predicting academic performance in each school year [24]. In a related paper [25], the first three years' performance of engineering students in Nigeria has been analysed using their graduation results. Six data mining algorithms were used, and the algorithm with the higher accuracy was selected.

Concerning ethical issues, the ethical issues in data mining are considered in [26]. Educational data mining is now seen at all levels of education and in all aspects of learning. The individual's privacy and inventiveness hold ethical implications for minors and the vulnerable in society. This was done by analysing a paper written by the U.S Department of Education. In [26], a review of data mining in education is done. From graduate theses and articles reviewed in Turkey, the data analysis techniques, sampling, nature of data and the data sources were considered. Variables for the study and the results were also supposed to know the target in educational data mining. It was realized that artificial neural networks are the most common educational data mining technique in solving academic problems. It is seen that students' achievement is the target in all the papers, and learners at the higher level are targeted. SPSS was seen as the recommended tool for data analysis. In a survey [27], a comprehensive review of educational data mining is done with associated terminologies such as learning analytics, data-driven decision-making, big data in education, educational data science etc. The main publications and findings regarding the objectives and the methodologies were considered. The future trend in the research area is also noted. In a blended learning course, the learning behaviours of learners were analysed. With experimental data from two classes in python programming, the prediction of higher-risk learners was accurately made. This is possible with data from the educational environment. Machine learning and symmetric-based learning approaches were used in the model to help in the prediction of students' behaviours. The validity of learning and the testing of learning styles theory using data mining can be seen in [27]. Evidence of learners' learning styles concerning the learning materials and the teaching techniques adopted was analysed using the logs of the learners and the instructors in a learning management system using the index of learning styles. However, the research proved no correlation between the learning style and the learning materials used. In [28], the evaluation of assessment questions for learners has been modelled using classification as a data mining technique. This analyses the questions based on bloom's taxonomy of assessment. Commonly, educational research is based on surveys, interviews and classroom observations as used in [29] to identify education challenges. In [30], classification in data mining can be used to gather detailed information on a problem. Proper categorization of schools into urban and rural gives more transparent information for policy interventions [31].

3. Methods

This is exploratory research. It makes use of both quantitative and qualitative data from the Ministry's EMIS reports in Ghana. The data is from a population of 216 districts in 10 regions. It is secondary data extracted from the year 2000 to 2019 academic years. EMIS conducts an annual census in basic schools to gather information on access, quality and education management. The data mining tool used in the analysis of the data is RapidMiner studio 9.10. Classification and clustering as data mining techniques are employed in grouping the data into rural and urban classes. Urban schools were seen as schools in communities which has a population of 5000 people or more. Rural schools were seen as schools in communities with a population of fewer than 5000 people. This is further grouped into clusters in the form of well-endowed and deprived regarding the challenges, accurate interventions needed and the solutions.

The qualitative data from reports were also extracted and presented in quotations where the information was quoted from the reports. It has been used in coming out with the findings of the quality of education in the country. The methodology used is technical and innovative as opposed to the methodology used in the available research in the area. Research works on BE in Ghana are presented in table 1.

Table 1. Research works in Basic Education in Ghana.

S/N	Ref.	Year	Topic/Aspect	Research Techniques used
1	[2]	2015	Examining the prospects and challenges of recent provisions	Purposive target interventions
2	[3]	2019	Basic Education in Ghana: Success, challenges and way forward	Survey analysis in graphs and tables
3	[6]	2007	Prospects and Challenges of the School Performance Improvement Plan	Percentages, charts and graphs
4	[7]	2015	Special and inclusive education in Ghana: Status and progress, challenges and implications	Case study
5	[5]	2016	Access to Basic Education in Ghana: The Evidence and the Issues	Case study
6	[9]	2018	Basic Statistics and Planning Parameters for Basic Education in Ghana	Charts, graphs and percentages.
7	[29]	2011	Education in Ghana – status and challenges	Charts, percentages and tables
8	[11]	2021	Enhancing Efficiency in Basic Educational Management using Data Mining Techniques	Classification and clustering

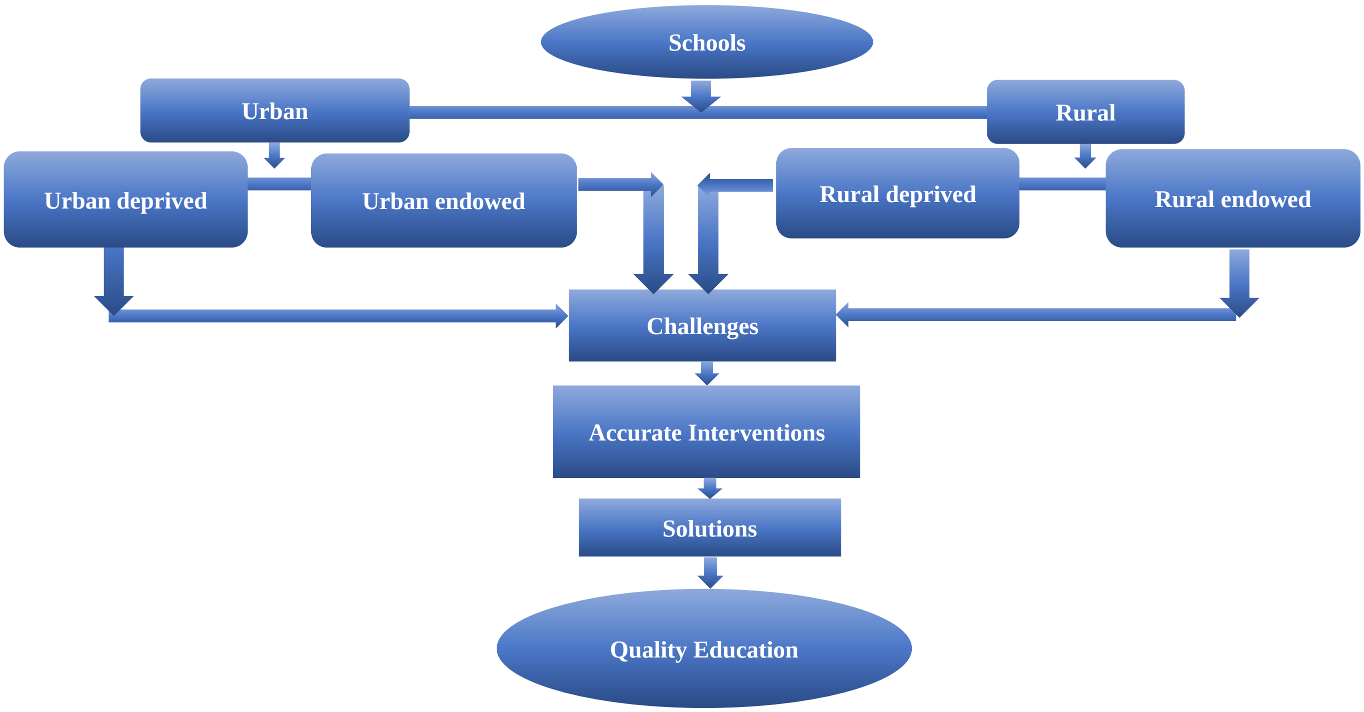

Fig. 1 Flowchart for the Problem Identified

3.1 Data Collection

The data used is from the MoE. Schools supplied the data to the ministry through the Ghana Education Service. This data is collected from an annual census of schools to know the available resources in schools and the needs of the schools. It is a legal requirement for every school to provide the data. The ministry summarizes this data, producing national, regional, district and school reports. Qualitative data augment the quantitative data from the reports to enhance the authenticity of the information and the datasets. For deep mining, a training dataset from one region was used.

3.2 Data Preparation

The data were filtered and normalized to enhance the integrity and efficient performance of the model. Null values were removed, and errors in data entry resulting in multiple entries in a cell were also removed. The incomplete data, which is not meaningful, were removed since they were irrelevant. This was done by visiting the individual variables which constituted the columns of the data. It made it more suitable for the use of various classification algorithms. As the data was extracted from reports, an amount of qualitative text was used. **Error! Reference source not found.** is the f lowchart for the work.

3.3 Feature Selection

There was a need for feature selection to maximize the model's efficiency and enhance the faster running of the selected classification algorithms. This helped in using only the relevant features in each of the algorithms to achieve the targets. Redundant variables were excluded in the selection to pave the way for more useful variables in other to avoid complexity. The various features that were selected are mainly on the locality of the schools, the level type, the names of the schools, the available enrolment by sex and class, the availability of teachers, the furniture (sitting and writing places) situation, teaching and learning materials, attendance and dropout.

3.4 Description of the Datasets

The data is made up of the information available in the various schools in each district of the region. It consists of the year of the census, the name of the region, the name of the district, the type of locality in which the school is (rural or urban), the level, the status (private or public), the EMIS code which is a unique ID for each school, name of the school, enrolment by age and sex for the various classes, teacher availability by sex for the various classes based on professionalism. Classroom availability, furniture available for the various classes, availability of reading materials, learners' attendance, and dropout. Specifically, the variables used are in two categories; educational resources and challenges. For the educational resources, the variables are; Sitting_Places (Sitting places), Writing_Places (writing places), Trs_Desk(Teachers Desk), Trs_Chair(Teachers Chair), Blackboard, Cupboard, Whiteboard, Eng_TxtBks(English Textbooks), Maths_ TxtBks (Mathematics Textbooks), Sci_ TxtBks(Science Textbooks), SStud_TxtBks (Social Studies Textbooks), MTrn(Male Trained Teachers), FTrn(Female Trained teachers) and ClassRms(Classrooms. All these are numeric variables. For the challenges, the variables are; MUntrn (Male Untrained teachers), FUntrn(Female Untrained teachers), MajRprs (classrooms needing major repairs), MinRprs (Classrooms needing minor repairs) and the dropout for all the years of study by sex.

3.5 Exploring the Data

Next, after knowing the data sets was the exploration of the data to have a feel of the likely direction and the results that may be realized. A summary of the data in the form of totals, averages, and minimum and maximum values was calculated for the numeric values. Simple graphs were also generated to have a pictorial view of the problem. The qualitative data, which is made up of text, was thoroughly proofread to identify errors and correct them. This was done to ensure the meaning of the statements was improved for better analysis.

3.6 Model Selection, Training and Evaluation

Nine classification algorithms were selected and used to analyze the training and test data sets to enhance maximum efficiency. As shown in table 1, the results of the algorithm with the highest efficiency are used. The algorithms used are; GLM, NB, LR, DL, DT, FLM, GBT, RF, and SVM. These algorithms are available in rapid miner studio 9.10. GBT was selected as the data mining algorithm for classification, which is suitable for the nature of the data. The training data set was used in training the model after the selection. It was done with all features available in the model using the classification algorithms. It was done to identify the most efficient model and use the features selected accurately. A comparison of the models is made to know the models with higher performance regarding the features selected. Six performance metrics were used in each model to compare with the others. They are the accuracy, receiver operating characteristics area under the curve (ROC-AUC), specificity, the F-measure, sensitivity, and precision. The result is presented in Table 2.

Table 2. Summary of the Performance of the Explored Algorithm

Metrics used	**Classification Algorithms Used**								
	NB	GLM	LR	FLM	DL	DT	RF	GBT	SVM
Accuracy	81.9	84.5	83.2	77.9	84.2	84.0	83.2	84.5	80.8
AUC	89.7	91.6	90.8	77.4	90.4	89.0	90.0	91.5	88.0
Precision	87.1	85.3	86.1	78.6	86.7	84.9	88.0	85.3	82.1
F-measure	87.5	89.0	87.8	85.0	88.5	88.6	87.5	89.0	86.6
Sensitivity	93.6	93.0	89.6	92.5	90.5	92.8	87.1	93.1	91.5
Specificity	57.5	66.8	70.1	47.7	71.5	65.9	74.9	66.4	58.4

Table 3. Gradient Boosted Tree Performance Results for the Classification of the Data

	true Urban	true Rural	class precision
pred. Urban	278	97	74.13%
pred. Rural	55	594	85.3%
class recall	83.48%	85.96%	

Table 4. Confusion Matrix for the classification

Criterion	Values	Standard Deviation
Accuracy	84.5	± 0.9%
AUC	91.5	± 0.9%
Precision	85.3	± 0.6%
F-measure	89.0	± 0.9%
Sensitivity	93.1	± 1.1%
Specificity	66.4	± 2.8%

4. Results and Discussion

The analyzed data is a case study of Ghanaian basic schools. The output is displayed in graphs and tables for easy interpretation. A summary of the performance of the classification algorithms used is presented as evidence for the choice of algorithm. The most efficient algorithm's detailed results are presented in figs 4,5, 6 and 7 and in tables 3 and 4.

4.1. Data Visualization

The data was made up of 3583 records with 64 attributes. Out of these, 2385 were records from rural schools, and the remaining 1153 were records from urban schools. 218 of these records were from preschool, 1380 were from kindergarten, 1419 were from primary, and the remaining 521 schools were Junior High Schools. The data was visualized in graphical form to have a clear picture of the nature of the problem. The resource availability in schools was viewed from the perspective of the locality type, and the year the school was created. This was to determine if urban schools are more resources than rural schools or vice versa. The year a school is created was also considered to know if resource distribution is affected by the year of creation (old schools or recently created schools). The school level was also supposed to know if that affected the problem. Figures 2 and 3 are graphical representations of the results.

4.1.1. Resource Availability and Distribution

Fig 2 shows the average resource availability in schools based on the locality. Sitting and writing places of the learners are the major resources identified for two classes, rural and urban. The two major resources are sitting places and writing places. In fig 3, the average resource availability by level is presented. The major resources are still the sitting and writing places.

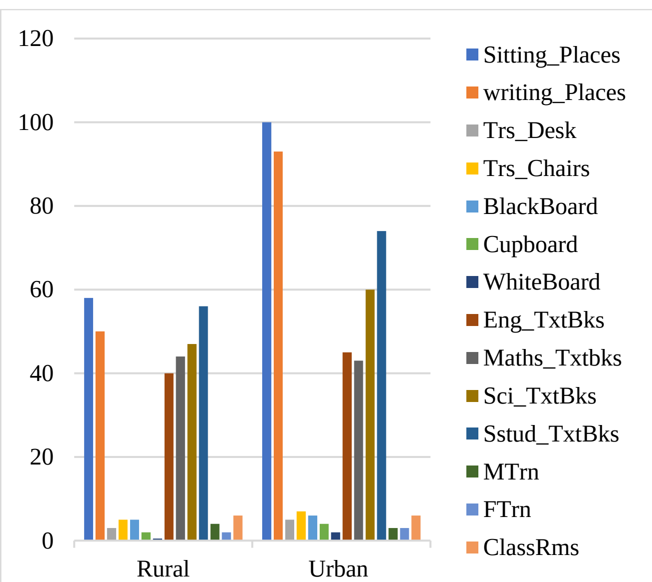

Fig. 2 Graphical Representation of the Available Resources in Schools by Locality Type

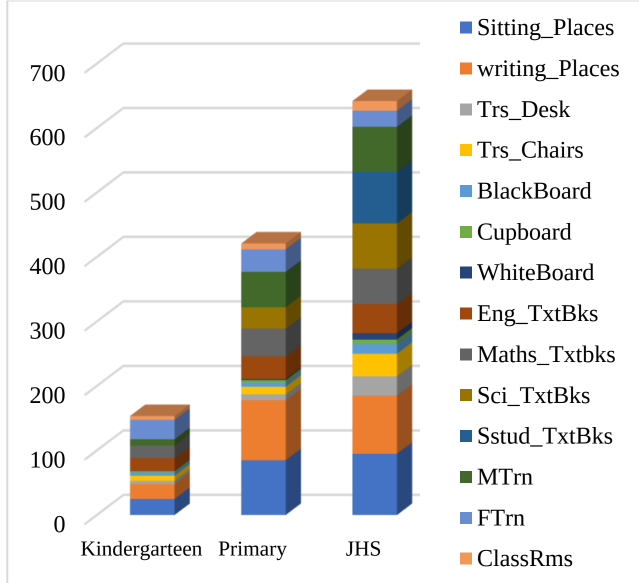

Fig. 3 Average Resource Availability in schools by the level of Education

4.1.2. Classification Algorithm Results

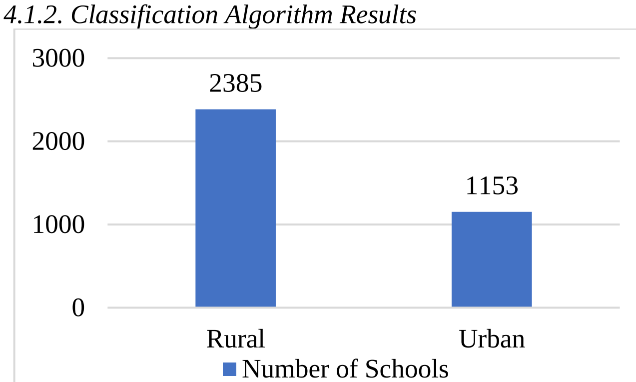

Fig. 4 Graphical Representation of the Rural and Urban Classes of the Schools.

4.2. Exploratory Results

The quest to get the most efficient algorithm to describe the data accurately lured the researchers to explore the classification algorithms available in Rapid Miner Studio 9.10. Nine classification algorithms were explored to analyze the data. These are; GLM, NB, LR, DL, DT, FLM, GBT, RF and SVM, as presented in table 2. From the overview of the data, the total enrolment of pupils at all levels of BE is 494579 for both urban and rural schools. Teachers handling these pupils are 16861 for both trained and untrained teachers. Classrooms to accommodate these learners totalled 11026. This means the pupil-classroom ratio is 45 pupils in a class. Sitting places for the learners totalled 231228 which means over 263351 of the learners do not have places to sit. This constitutes 53.25% of the learners. Writing places for the learners totalled 206157, serving only 41.68% of learners in school. The remaining 58.31% do not have places to do the writing.

Also, this means over 57194 learners sit to listen to the teacher but cannot write because there is no place for them. For the aspect of teacher sitting places and writing places, a total of 6408 desks and 9464 chairs are available to serve the 16861 teachers. This means more than 50% of teachers do not have access to desks and chairs in the classroom. Classroom learning materials such as blackboard and whiteboard are 9661 and 1911, respectively. This totalled 11572, sufficient for the 11026 classrooms if distributed equally. Only 4631 cupboards are available for the 11026 classrooms. This is woefully inadequate for storing the learning materials. For the aspect of reading materials, English language, mathematics, science, and social studies recorded 41002, 43798, 29670, and 18105, respectively. This is inadequate for the total enrolment of 494579 learners at all levels. Sitting places are the major resource in basic schools. This will therefore be used for the prediction of endowed and deprived schools.

4.2.1. Clustering

After the various classes of urban and rural schools were identified, the further grouping of the schools into endowed and well-endowed schools was necessary. This was to identify schools with relatively available resources. The initial grouping was done to know the generally endowed and deprived schools without considering rural or urban. Two clusters were identified in this; cluster_0 for the deprived schools and cluster_1 for the endowed schools by using the sitting places as primary resource determined. Using the k-Means clustering algorithm, Cluster_0 showed 3152 while cluster_1 showed 386. Averages were used to indicate the schools with sitting places available for the learners. This was done by comparing the available sitting places to the enrolment.

4.2.2. Clustering for Rural

For the class of rural schools, x-means clustering is used to group the schools into two major groups: rural endowed and rural deprived. Cluster 0 is for the rural deprived, and cluster 1 is for the rural endowed. Figure 5 shows the summary of the clustering. Sitting places were a major resource identified separating the rural endowed from the rural deprived.

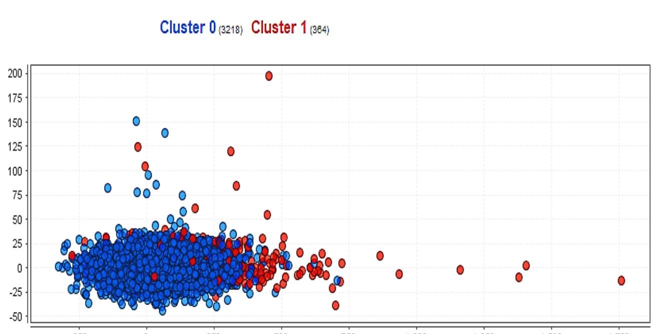

Fig. 5 Graphical Representation of general Clustering of the schools

Fig. 6 X-Means Centroid Chart for Rural Deprived and Endowed Schools.

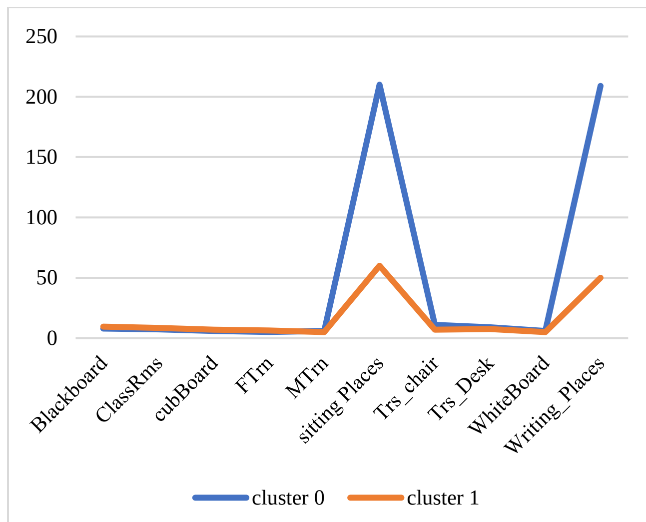

Fig. 7 x-means centroid chart for urban endowed and urban deprived schools.

4.2.3. Clustering of Urban Data

This data was also grouped into two major categories following the same criteria in clustering the rural schools. Two major classes (urban endowed and urban deprived) were identified—cluster 0 for the urban deprived and cluster 1 for the rural deprived. Figure 8 shows the summary of the results.

4.3. Challenges at the Various Levels of Education

Enrolment of children of school-going ages keeps increasing, but a significant number of these children are not in school. Resources available for those in schools are inadequate, especially in the rural deprived schools. Inadequate professional teachers, inadequate sitting and writing places, inadequate reading materials, etc., are the major challenges of BE. These challenges are noticed in every report on BE every year. For decades in collecting and analyzing data by the MoE, this problem is not solved, leading to low pass rates, which affect the quality of education. The poor quality affects the entrants to the next level of education. A report from a director-general has the information below.

"The basic education system is challenged in access and participation, which affect quality. As a result of the inconveniences learners go through due to unfriendly learning environments, girls' enrolment is very low in schools. The poor enrolment of girls affects the gender parity index, and the unfriendly school environment affects the learning outcomes. The inadequate sitting and writing places, inadequate reading materials and low teacher availability in schools affect the quality of pedagogy in primary schools. As Ghana abolished the payment of school fees, schools solely depend on the government for support, and there is low parental participation" The challenges of Senior High Schools in Ghana start from low enrolment due to inadequate space and low pass rates for the Basic School leavers. Educational resources are relatively lower at this level than at the Basic level. Access to senior high school education was very low in Ghana due to the high cost and low accessibility. The inability to absorb all the basic school graduates leads to dropout. However, the government of Ghana has now made senior high school education free in the country. Despite the intervention takes, the challenges have increased as more students are now enrolled to compete with the inadequate resources available in the insufficient senior high schools in the country. This has compelled the government to order senior high schools to run tracks and operate on a semester basis in other to enhance resource sharing. The sharing of resources has enhanced more utilization of seating and writing places, giving each student access to a seating and writing place in the school. The statement below is an extract from the reports

"There are redundant learning resources as a result of the change in curriculum and change in educational policies. Student-textbook ratios have reduced drastically since the 2011/12 academic year, reaching 0.5 in the 2016/17 academic year for math, and textbook production is often delayed. The challenge keeps appalling since enrolments increase yearly and learning materials are unavailable. The need to increase infrastructure to cater for learners is also a major problem since student congestion in the class is becoming a norm. This happens without a corresponding increase in the number of teachers and without building the capacity of already existing teachers. The effect of this on learning outcomes shows a poor pass rate. Rural deprived schools are the most affected." These institutions are not common in Ghana, and the enrolment is the lowest compared to other levels. Until recently, it was not attractive due to many students' misconceptions about it. It was perceived that students who do not pass well attend such schools. The requirement of practical based training and the use of expensive tools discouraged some students from attending such schools.

However, unemployment challenges are now making it attractive as students have realized the need to have the skills to be self-employed. Ghana's government has also attempted to increase access by providing more schools and including some of the technical and vocational institutions in the Free Senior High School Programme (FSHSP). Below is a comment from the reports. "*Access in technical and vocational schools has increased as a result of government commitment to make this level attractive to learners; infrastructure has been increased to cater for more students, and laboratories are provided for practical learning. This, however, is still constrained by low public perceptions: there is a lukewarm attitude towards technical schools. Female participation in this level is very low, and they are usually clustered in handy works such as dressmaking, hairdressing etc. Modification of the curriculum of this level is necessary to produce graduates who will solve problems in the Ghanaian economy.*

The universities in Ghana are not adequate to absorb all graduates of the Senior High Schools and the Vocational and Technical institutions. Public regular universities are currently nine (9), and ten (10) polytechnics have been upgraded to technical universities. Distance education and sandwich programs have been mounted in all public universities. All these measures are to increase access to tertiary education. Public universities' attempt to increase access has resulted in a larger number of students in lecture halls, making it difficult for lecturers to use the appropriate teaching methodologies to teach. *"Regarding quality, the recommended student-teacher ratio and the student-classroom ratio are appalling in the public schools. Except for the sciences, there is congestion in the lecture halls, and a lecturer is compelled to use the lecture delivery method to meet the content and time. Even with this challenge, less than 40% of the lecturers have the minimum qualification to teach at the tertiary level. This is worst in rural deprived universities. On average, a lecturer meets only 0.5 required publications for a year. This means research is a major problem. Also, no Ghanaian university has ever been ranked best in the African continent."* It is clear from the reports that quality education is a major issue in Ghana. This, however, is cyclical rotating among all levels. Success in the form of quality basic education is leveraged to achieve the success of quality pre-tertiary education, which is also leveraged to achieve the success of quality tertiary education. Qualified tertiary education graduates are the teachers to improve the quality of teaching and learning in basic schools. Fig 8 is a diagram representing the relationship between the levels of education in achieving quality education. Fig 9 is a neural network diagram for the challenges and solutions to enhance quality education.

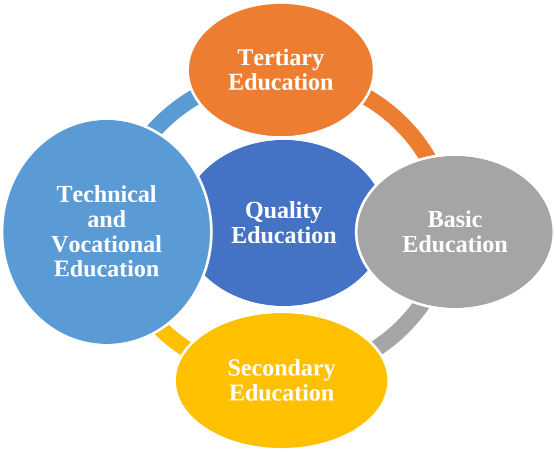

Fig. 8 The Interoperability of Quality Education among Different Levels.

4.4. Critical Issues

It is clear from the results of the analysis that educational resources are not fairly distributed among rural and urban schools. Further discrimination of both rural and urban schools creates segregation of endowed and deprived schools. This clearly shows that not all schools have the available resources for effective teaching and learning.

National assessment of learners for promotion to the next level, however, does not discriminate among the schools putting the deprived schools at a disadvantage. Significant among the resources is the sitting and writing places. This is woefully inadequate as most schools do not have enough of these resources to cater to all learners. An unfriendly learning environment is an effect, and poor performance will be the result. Rural schools are the most affected by discrimination, and deprived schools are the worst. For two decades and beyond, these challenges are affecting the quality of education in the country. As indicated in [4], many governments have come and gone, but the problems persist. Ghana needs policy stability to help in improving the education sector. Fluctuations in educational policies confuse the system and affect the allocation of resources to schools. The political transition of leadership in Ghana has been very smooth since Ghana moved to the fourth republic. However, this peaceful transition of power does not go with stable educational policies.

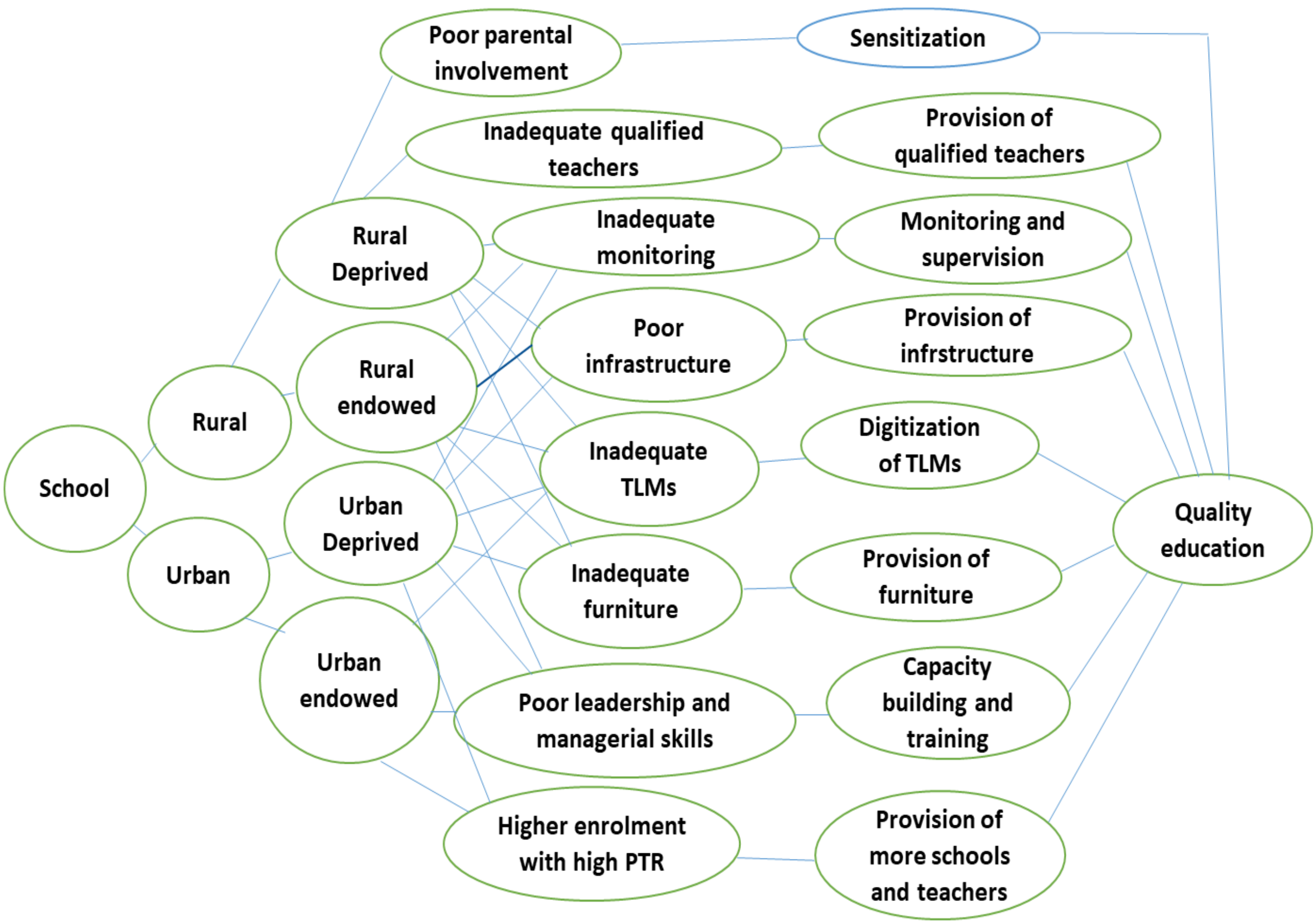

Fig. 9 Proposed Neural Network Model for Managing Basic Education

Urban schools in Ghana are densely populated with higher enrolments making it too competitive for the resources available. The congestion in the classrooms puts pressure on the infrastructure and the available teaching and learning materials. Multiple streams characterize some urban schools for a level due to the higher enrolments. In some cases, the schools are compelled to run shift systems. Though urban schools have relatively higher resources compared to rural schools, the higher enrolments result in a higher pupil-teacher ratio and pupil-classroom ratio as well as a pupil-textbook ratio which negatively affects quality teaching and learning.

Low enrolments characterize rural schools due to the lower population of the areas. Poor infrastructural facilities and Teaching and learning materials negatively affect the quality of teaching and learning. Qualified teachers in rural schools are relatively few as compared to urban schools. This creates room for most unqualified teachers in rural schools since they lack qualified teachers. Due to the lower populations, pupils and teachers are compelled to travel a reasonable distance to schools closer to them. This puts stress on both the teacher and the pupils. Using a case study as in [2], [3] does not identify the challenges and needs of BE in Ghana. With the use of classification algorithms, the attributes of educational data which spell out the challenges and solutions to the education problem are accurately identified. Various levels of basic education with their associated needs have been identified. Data mining techniques have been used to identify and train a model which can assist in predicting and planning for the educational resources in a country. The feature selection has been used to accurately pick the relevant attributes that can improve a country's education quality. The results have been presented in the form of tables and graphs regarding the various levels of education and the challenges and needs associated with them.

4.5. Theoretical Contribution and Implications

Already existing literature on identifying education challenges and getting sustainable solutions did not use appropriate tools and techniques to address the problem. The use of tables and graphs with illustrations and reporting in the form of a case study to identify the challenges of education did not help in accurately identifying the problem as indicated in [2], [3], [4], [5], [6], [7], [9], [11] and [29]. The novelty of using data mining techniques to identify the challenges of basic education and get appropriate interventions to address the problem is a major contribution to managing BE. This can help improve the entire education sector, which will help improve productivity in all sectors of a country since education is the foundation for growth and development. With data mining techniques, the trend of data from the education sector can be analyzed to identify the pattern. This can help leverage BE's success by minimizing cost and maximizing the judicious use of resources.

4.5.1. Implications for Practice

Using just one algorithm to analyze data and make predictions does not enhance accuracy in model training. Different algorithms should be used to analyze the same data to come out with the most efficient one suitable for the data. Overfitting of data, inaccuracies and low speed in computation can be eliminated when different algorithms are used to come out with the best algorithm. There is no ideal classification algorithm; the data's nature dictates the classification algorithm that should be used. The interpretation of the algorithm's results also strongly depends on the nature of the data.

5. Conclusion

Despite the importance attached to education in a country, data mining in educational management is uncommon, especially in the management of BE. The exponential rate at which new schools are established to cater to the population of the school-going ages calls for the use of machine learning or data mining to analyse educational data. The need for a model to help in predicting and planning the resources of BE is also a necessity. But as the management of BE is characterized by massive personal interest, the readiness of educational management to accept proposed models to predict and plan for the utilization of educational resources is questionable. In addition, education policies are fluctuating in many countries due to changes in political power. The policy instability will affect the quality of education as the flow of resources may be affected by the politics in a country.

References

[1] Jing Rao, and Jingzhong Ye, “From a Virtuous Cycle of Rural-Urban Education to Urban-Oriented Rural Basic Education in China: An Explanation of the Failure of China’s Rural School Mapping Adjustment Policy,” *Journal of Rural Studies*, vol. 47, pp. 601–611, 2016. *Crossref*, http://doi.org/10.1016/j.jrurstud.2016.07.005

[2] Ziblim Abukari, Ahmed Bawa Kuyini, and Abdulai Kuyini Mohammed, “Education and Health Care Policies in Ghana: Examining the Prospects and Challenges of Recent Provisions,” *SAGE Open*, vol. 5, no. 4, 2015. *Crossref*, http://doi.org/10.1177/2158244015611454

[3] D. K. Mereku, “Sixty Years of Teacher Education in Ghana: Successes, Challenges and the Way Forward,” *African Journal of Educational Studies in Mathematics and Sciences*, vol. 15, no. 2, pp. 69–75, 2019. *Crossref*, http://doi.org/10.4314/ajesms.v15i2.6

[4] K. Asare and P. Osei-Kuffour, “Briefing Paper,” vol. 18. no. 2, 2021.

[5] Kwame Akyeampong et al., "*Access to Basic Education in Ghana : The Evidence and the Issues*," The Consortium for Educational Access, Transitions and Equity (CREATE), 2007.

[6] Joseph Ghartey Ampiah et al., "*Improving Quality Basic Education in Ghana : Prospects and Challenges of the School Performance Improvement Plan,*" University of Cape Coast Institutional Repository, pp. 73–98, 2007.

[7] Lawrence KofiAmetepee, and Dimitris Anastasiou, "Special and Inclusive Education in Ghana: Status and Progress, Challenges and Implications," *International Journal of Educational Development*, vol. 41, pp. 143–152, 2015. *Crossref,* http://doi.org/10.1016/j.ijedudev.2015.02.007

[8] Kwame Akyeampong, "Public-Private Partnership in the Provision of Basic Education in Ghana: Challenges and Choices," *Compare: A Journal of Comparative and International Education*, vol. 39, no. 2, pp. 135–149, 2009. *Crossref,* http://doi.org/10.1080/03057920902750368

[9] E. June, "*Report on Basic Statistics and Planning Parameters for Basic Education in Ghana*," Ministry of Education, Republic of Ghana 2018.

[10] E. June, "*Report on Basic Statistics and Planning Parameters for Basic Education in Ghana*," Ministry of Education, Republic of Ghana, 2014.

[11] Fuseini Inusah et al., "Expert System in Enhancing Efficiency in Basic Educational Management using Data Mining Techniques," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, pp. 427-434, 2021. *Crossref,* http://doi.org/10.14569/IJACSA.2021.0121148

[12] Fuseini Inusah, and Anokye Acheampong Amponsah, "An Expert System to Assist Businesses in Financial Decision Making to Enhance Efficiency," *International Journal of Computer Applications*, vol. 181, no. 7, pp. 32-39, 2018.

[13] AlejandroPeña-Ayala, "Educational Data Mining: A Survey and a Data Mining-Based Analysis of Recent Works," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1432–1462, 2014. *Crossref,* http://doi.org/10.1016/j.eswa.2013.08.042

[14] Felipe Azevedo Pinage, Eulanda Miranda dos Santos, and João Manuel Portela da Gama, "Classification Systems in Dynamic Environments: An Overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 6, no. 5, pp. 156–166, 2016. *Crossref,* http://doi.org/10.1002/widm.1184

[15] Marco A.F.Pimentel et al., "A Review of Novelty Detection," *Signal Processing*, vol. 99, pp. 215–249, 2014. *Crossref,* http://doi.org/10.1016/j.sigpro.2013.12.026

[16] Michail Papathomas, amd Sylvia Richardson, "Exploring Dependence Between Categorical Variables: Benefits and Limitations of Using Variable Selection within Bayesian Clustering in Relation to Log-Linear Modelling with Interaction Terms," *Journal of Statistical Planning and Inference*, vol. 173, pp. 47–63, 2016. *Crossref,* https://doi.org/10.1016/j.jspi.2016.01.002

[17] Mohamed Osman Hegazi, and Mazahir Abdelrhman Abugroon, "The State of the Art on Educational Data Mining in Higher Education," *International Journal of Computer Trends and Technology,* vol. 31, no. 1, pp. 46-56, 2016.

[18] Hirley Alves, and Onel Alcaraz Lopez, "Appendix C: Clustering Algorithms," *in Wireless RF Energy Transfer in the Massive IoT Era: Towards Sustainable Zero-energy Networks, IEEE,* pp. 198-201, 2022. *Crossref,* https://doi.org/10.1002/9781119718703.app3

[19] Mudasir Ashraf, Majid Zaman, and Muheet Ahmed, "An Intelligent Prediction System for Educational Data Mining Based on Ensemble and Filtering Approaches," *Procedia Computer Science*, vol. 167, pp. 1471–1483, 2020. *Crossref,* https://doi.org/10.1016/j.procs.2020.03.358.

[20] Phauk Sokkhey, and Takeo Okazaki, "Developing Web-Based Support Systems for Predicting Poor-Performing Students Using Educational Data Mining Techniques," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, pp. 23–32, 2020. *Crossref,* https://doi.org/10.14569/IJACSA.2020.0110704

[21] S. Križanić, "Educational Data Mining Using Cluster Analysis and Decision Tree Technique: A Case Study," *International Journal of Engineering Business Management*, vol. 12, 2020. *Crossref,* https://doi.org/10.1177/1847979020908675

[22] Johnsymol Joy, "Overview of Different Data Clustering Algorithms for Static and Dynamic Data Sets," *SSRG International Journal of Computer Science and Engineering,* vol. 5, no. 3, pp. 1-3, 2018. *Crossref,* https://doi.org/10.14445/23488387/IJCSE-V5I3P101

[23] Luis Naito Mendes Bezerra, and Márcia Terra Silva, "Educational Data Mining Applied to a Massive Course," *International Journal of Distance Education Technologies,* vol. 18, no. 4, pp. 17–30, 2020. *Crossref,* https://doi.org/10.4018/IJDET.2020100102

[24] Eduardo Fernandes et al., "Educational Data Mining: Predictive Analysis of Academic Performance of Public School Students in the Capital of Brazil," *Journal of Business Research*, vol. 94, pp. 335–343, 2019. *Crossref,* https://doi.org/10.1016/j.jbusres.2018.02.012

[25] Aderibigbe Israel Adekitan, and Odunayo Salau, "The Impact of Engineering Students' Performance in the First Three Years on their Graduation Result Using Educational Data Mining," *Heliyon*, vol. 5, no. 2, p. e01250, 2019. *Crossref,* https://doi.org/10.1016/j.heliyon.2019.e01250

[26] Isak Potgieter, "Privacy Concerns in Educational Data Mining and Learning Analytics," *The International Review of Information Ethics*, vol. 28, pp. 1–6, 2020. *Crossref,* https://doi.org/10.29173/irie384

[27] Cristobal Romero, and Sebastian Ventura, "Educational Data Mining and Learning Analytics: An Updated Survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, p. e1355, 2020. *Crossref*, https://doi.org/10.1002/widm.1355

[28] Anwar Ali Yahya, "Swarm Intelligence-Based Approach for Educational Data Classification," *Journal of King Saud University - Computer and Information Sciences*, vol. 31, no. 1, pp. 35–51, 2019. *Crossref*, https://doi.org/10.1016/j.jksuci.2017.08.002

[29] Charles Aheto-Tsegah, "Education in Ghana – Status and Challenges," *Commonwealth Education Partnerships*, pp. 27–30, 2011. [Online]. Available: http://www.cedol.org/wp-content/uploads/2012/02/Charles-Aheto-Tsegah-article.pdf

[30] R.Nirmala Devi, and Anupriya Sharma, "Effective Active Course Learning Method for Education Data Using Feature Base Classification Model," *International Journal of Engineering Trends and Technology*, vol. 67, no. 10, 2019.

[31] Ana Carolina Lima de Souza, Tiago Bittencourt Nazaré, and Douglas Barbosa, "Localization of the Elementary EducationSchools: an Application to Urban Municipal Schools of Cataguases/MG – Brazil," *International Journal of Engineering Trends and Technology*, vol. 63, no. 1, pp. 1-5, 2018.

[32] Xinlei Wanga, Soohyun Ahnb, and Johan Lim, "Unbalanced Ranked Set Sampling in Cluster Randomized Studies," *Journal of Statistical Planning and Inference*, vol. 187, pp. 1–16, 2017. *Crossref*, https://doi.org/10.1016/j.jspi.2017.02.005

[33] Aberbach Hicham et al., "A Survey on Educational Data Mining [2014-2019]," *2020 International Conference on Intelligent Systems and Computer Vision, ISCV 2020*, pp. 1-6, 2020. *Crossref*, https://doi.org/10.1109/ISCV49265.2020.9204013