

# Evidence of Students' Academic Performance at the Federal College of Education Asaba Nigeria: Mining Education Data

Arnold Adimabua Ojugo <sup>a,1,\*</sup>, Christopher Chukwufunaya Odiakaose <sup>b,2</sup>, Frances Emordi <sup>c,3</sup>, Rita Erhovwo Ako <sup>a,4</sup>, Winifred Adigwe <sup>d,5</sup>, Kizito Eluemonor Anazia <sup>e,6</sup>, and Victor Geteloma <sup>a,7</sup>

<sup>a</sup> Department of Computer Science, Federal University of Petroleum Resources  
PMB 1221 FUPRE road, Ugbomro, Effurun 330102, Nigeria

<sup>b</sup> Department of Computer Science, Faculty of Information Technology, Dennis Osadebay University  
Bonsaac, Pantor drive, Anwai-Asaba 320006, Nigeria

<sup>c</sup> Department of CyberSecurity, Faculty of Information Technology, Dennis Osadebay University  
Bonsaac, Pantor drive, Anwai-Asaba 320006, Nigeria

<sup>d</sup> Department of Computer Science, Faculty of Information Technology, University of Science and Technology Ozoro  
Kwale Rd, Ozoro 334113, Nigeria

<sup>e</sup> Department of Information Technology, Faculty of Information Technology, University of Science and Technology  
Ozoro

Kwale Rd, Ozoro 334113, Nigeria

<sup>1</sup>ojugo.arnold@fupre.edu.ng\*; <sup>2</sup>osegalaxy@gmail.com; <sup>3</sup>frances.emordi@dou.edu.ng; <sup>4</sup>ochukorita2@gmail.com;  
<sup>5</sup>adigwew@dsust.edu.ng; <sup>6</sup>anaziake@dsust.edu.ng; <sup>7</sup>vochuko@gmail.com

\* corresponding author

## ARTICLE INFO

## ABSTRACT

### Article history:

Received 04 October 2022  
Revised 09 October 2023  
Accepted 14 October 2023  
Published online 19 October 2023

### Keywords:

Academic Performance  
Bayesian Network  
Educational Data  
Decision Tree  
Summative Testing

One main objective of higher education is to provide quality education to its students. One way to achieve the highest level of quality in the higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students, and prediction about students' performance. The knowledge is hidden among the educational data set and is extractable through data mining techniques. The present paper is designed to justify the capabilities of data mining techniques in the context of higher education by offering a data mining model for the higher education system in the university. In this research, the classification task is used to evaluate student's performance, and as many approaches are used for data classification, the decision tree method is used here. By this, we extract data that describes students' summative performance at semester's end, helps to identify the dropouts and students who need special attention, and allows the teacher to provide appropriate advising/counseling.

This is an open access article under the CC BY-SA license  
(<https://creativecommons.org/licenses/by-sa/4.0/>).

## I. Introduction

The advent of data technology in various fields has led to massive volumes of data in various forms like files, audio, videos, images, and lots of new data formats [1][2]. Data from diverse applications requires a correct method of extracting knowledge from large repositories for better decision-making [3]. Knowledge discovery aims at birthing valuable, meaningful data via a collection of knowledge [4][5]. Knowledge mining uses various methods and algorithms to extract various forms of data. Data processing and mining for knowledge discovery tools have since recorded tremendous success in their impact [6][7][8], and have become an essential facet in various organizations [9][10][11][12]. Data processing techniques are introduced into new fields of statistics, databases, machine learning, pattern reorganization, AI, and Computation competencies.

There are growing study interests in using educational data mining. This recently evolving field, called educational data mining, concerns developing approaches that discover knowledge from data originating from educational environments [13][14][15][16]. Educational data mining uses techniques like Decision Trees, Neural Networks, Naïve Bayes, and K-nearest neighbors [17]. These techniques reveal many sorts of knowledge, like association rules, classifications, and clustering [18][19]. The

revealed knowledge is used in prediction about the enrolment of scholars during a particular course, alienation of traditional classroom teaching model, detection of unfair means utilized in online examination, detection of abnormal values within the result sheets of the scholars, prediction about students' performance then on [20][21][22][23].

The study uses data mining methodologies to investigate students' performance within the various courses. Data mining offers many tasks to investigate student performance, and for such tasks in classification [24] – we seek to study student's performance using decision tree classification. Data such as class tests, attendance, assignment marks, and examination score(s) were collected and used to predict the performance at the top of the semester.

## II. Methods

Data Mining is often utilized in the tutorial field to reinforce our understanding of the training process to specialize in identifying, extracting, and evaluating variables associated with the training process of scholars as described by many scholars. Mining in an academic environment is named educational data mining. Data mining in education may be a recent research field and this area of research area is gaining popularity due to its potential for educational institutes.

Shiokawa et al. [25] describe data mining transactions as one that permits the users to research data from different dimensions, categorize it, and summarize the relationships identified during the mining process [25]. The study was conducted on student performance by selecting 600 students from different colleges at Awadh University, Faizabad, India. Employing Bayes Classification on category, language, and background qualification, it had been found whether newcomer students will perform or not [26].

Ahmad et al. [27] conducted a study on student performance by selecting 300 students (225 males, 75 females) from a gaggle of schools affiliated to Punjab University of Pakistan. The hypothesis stated as "Student's attitude towards attendance in school, hours spent during a study on day to day after college, student's family income, students' mother's age, and mother's education are significantly related with student performance" was framed. Employing simple rectilinear regression analysis found that factors like the mother's education and the student's family income were highly correlated with the student's academic performance.

Brindlmayer [28] conducted a performance study on 400 students comprising 200 boys and 200 girls selected from the senior lyceum of Aligarh Muslim University, Aligarh, India, with the most objective to determine the prognostic value of various measures of cognition, personality, and demographic variables for fulfillment at higher secondary level in science stream. The choice was supported by the cluster sampling technique, during which the whole population of interest was divided into groups or clusters, and a random sample of those clusters was selected for further analyses. It had been found that girls with high socio-economic status had relatively higher academic achievement within the science stream, and boys with low socio-economic status generally had relatively higher academic achievement.

Nguyen et al. [29] gave a case study using student data to research their learning behavior to predict the results and warn students in danger before their final exams. They applied a decision (choice) tree model to predict the ultimate grade of scholars who studied the C++ course at Yarmouk University, Jordan, 2015. The 3-classification methods were used, namely ID3, C4.5, and the Naive Bayes. Their results indicated that the choice Tree model had better predictions than others. Also, Nilam [30] conducted a study on student performance by selecting 60 students from a degree college of Awadh University in India. Through association rule, they find the interesting in opting class teaching language. He describes using the k-means clustering algorithm to predict student's learning activities. The knowledge generated after implementing the info-mining technique could also be helpful for a teacher and college kids.

Chen [31], in his study on private tutoring and its implications, observed that the share of scholars receiving private tutoring in India was relatively above in Malaysia, Singapore, Japan, China, and Sri Lanka. It had also been observed that there was an enhancement of educational performance with the

intensity of personal tutoring, and this variation of intensity of personal tutoring depends on the collective factor, namely socio-economic conditions. Haiping et al. [32] conducted a study on student performance by selecting 300 students from 5 different degree colleges conducting the BCA (Bachelor of Computer Application) course. Through the Bayesian classification method on 17 attributes [33][34][35][36], it was found that factors like students' grades within the senior secondary exam, living location, medium of teaching, mother's qualification, student other habits, family annual income, and student's family status were highly correlated with the scholar academic performance.

Data mining, also called Knowledge Discovery in Database (KDD), often refers to digging out [37][38][39] or "mining" knowledge from large amounts of data. Data mining techniques are wont to operate vast volumes of data to get hidden patterns and relationships helpful in deciding [40][41][42]. While data mining and knowledge discovery within the database are frequently treated as synonyms, data mining is a component of the knowledge discovery process [43][44][45]. The sequences of steps identified in extracting knowledge from data are shown in Figure 1.

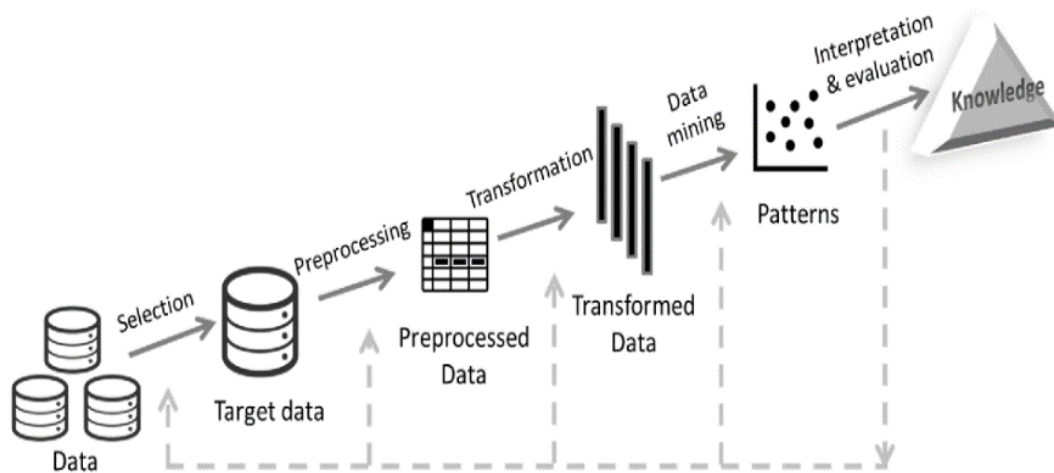


Fig. 1. The steps of extracting knowledge from data

Numerous heuristic techniques, such as classification, clustering, regression, neural networks, association rules, decision trees, and genetic algorithms, have been successfully used for database knowledge discovery. These techniques and procedures in data mining need to be briefly mentioned to understand better [46][47][48][49].

Regression techniques are often adapted for prediction. Regression or multivariate analyses are often used to model the connection between one or more independent variables and dependent variables. In data mining, independent variables are attributes already known, and response variables are what we would like to predict. Unfortunately, many real-world problems are not predictive [50][51][52][53]. Prediction is a declaration of something self-evident, a task that can be assumed as the basis for argument. Thus, more complex techniques are used (e.g., logistic regression, decision trees, or neural nets) to forecast future values. An equivalent model type can often be used for both regression and classification. For instance, the CART (Classification and Regression Trees) decision tree algorithms are often built to classify categorical response variables and to forecast continuous response variables. Neural networks can also create classification and regression models [21][54].

Classification is the most ordinarily applied data processing technique, which employs a group of pre-classified examples to develop a model to classify the population of records at large. This approach frequently employs decision trees or neural network-based classification algorithms. The info classification process involves Learning and classification. In Learning, the classification algorithm analyzes the training data [55]. In classification test data are wont to estimate the accuracy of the classification rules. If the accuracy is suitable, the principles are often applied to the new data tuples [23]. The classifier-training algorithm uses these pre-classified examples to determine the parameters required for correct discrimination. The algorithm encodes these feats into a model/classifier [26][56][57][58].

The decision tree is a tree-shaped structure representing selection sets. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees [59][60][61].

Clustering is often said because of the identification of comparable classes of objects. Using clustering techniques, we will further identify dense and sparse regions in object space and may discover overall distribution patterns and correlations among data attributes [62][63][64]. The classification approach can also be used to distinguish groups or classes of an object effectively but becomes costly, so clustering is often used as a preprocessing approach for attribute subset selection and classification [65][66][67].

Association and correlation usually seek out frequent item set findings among large data sets. This finding helps businesses form certain decisions, like catalog design, cross-marketing, and customer shopping behavior analysis. Association Rule algorithms have to be ready to generate rules confidently values but one. Also, the possible association rules for a given dataset are usually extensive, and many principles are usually of little (if any) value [54].

A neural network may be a set of connected input/output units, and every connection features a weight present with it. At its training phase, the network learns by adjusting weights to be ready to predict the suitable class labels of the input tuples. Neural networks can derive meaning from complicated or imprecise data and may want to extract patterns and detect trends that are too complex to be noticed by humans or other computer techniques [25]. These are compatible with continuous-valued inputs and outputs. Neural networks are best at identifying patterns or trends in data and are compatible with prediction or forecasting needs [68][69][70][71].

A technique that classifies each record during a dataset supported a mixture of the classes of the k record(s) most almost like it during a historical dataset (where k is bigger than or adequate to 1). Sometimes called the k-nearest neighbor technique [72][73].

#### A. Technical Experimental Framework

Today, a student's academic performance is determined by the internal assessment (formative tests) and end-of-the-semester (summative tests) examinations. The teacher administers the internal assessment based on students' performance in educational activities like class tests, seminars, assignments, general proficiency, attendance, and lab work. The scholar scores the end-of-semester examination within the semester examination. Each student must get minimum marks to pass a semester in internal and end semester examination.

The dataset used in this study was obtained from the Federal College of Education (Technical) Asaba, Delta State on the sampling method of the Post Graduate Diploma, of course, PDE-Technical Education (Post Graduate Diploma in Education, Technical Education Option) from session 2017 to 2020. Initially, the dimension of the info is 50. During this step, data stored in several tables was joined during a single table after joining process errors were removed.

In this step, only those fields were selected which were required for data processing. A couple of derived variables were selected. In contrast, some knowledge of the variables was extracted from the database. Table 1 gives all the predictor and response variables derived from the database for reference.

Table 1. Student related variables

Items	Description	Possible Values
CGP	Cumulative Grade Point	{Distinction >4.50, Credit <4.50 & >3.50, Merit <3.50 & >2.50, Fail <2.50}
TP	Teaching Practice	{"A" >70, "B" <70 & >60, "C" <60 & >50, "F" <50}
ASS	Assignment	{Yes, No}
GA	General Aptitude	{Yes, No}
ATT	Attendance	{Good, Average, Poor}
EP	Education Project	{"A" >70, "B" <70 & >60, "C" <60 & >50, "F" <50}
CGPA	Cummulatibe Grade Point Average	{Distinction >4.50, Credit <4.50 & >3.50, Merit <3.50 & > 2.50, Fail <2.50}

- The domain values for a few of the variables were defined for this investigation as follows:
- CGP – Cumulative grade point obtained at the end of semester examinations. CGP is split into four classes: Distinction >4.50, Credit <4.50 & >3.50, Merit <3.50 & >2.50, Fail <2.50.
  - TP – Teaching practice performance obtained in the final semester: Teaching Practice programs are organized to see the performance of scholars in teaching as a profession. Teaching practice is evaluated into four classes: “A” >70, “B” <70 & >60, “C” <60 & >50, “F” <50.
  - ASS – Assignment performance. In each semester, two assignments are given to students by each teacher. Assignment performance is split into two classes: Yes – student-submitted assignment, No – Student not submitted assignment.
  - GA - General Aptitude. Like seminars, in each semester, general proficiency tests are organized. The General Proficiency test is split into two classes: Yes – student generally participated proficiency, No – Student not generally participated proficiency.
  - ATT – Attendance of Student. A minimum of 70% attendance is compulsory to participate in the End Semester Examination. However, albeit in exceptional cases, low-attendance students also participate in the End Semester Examination for genuine reasons. Attendance is split into three classes: Poor - <60%, Average - > 60%, and <80%, Good - >80%.
  - EP – Education Project. The education project is split into two classes: Yes – student completed education project, No – student not completed education project. Education project as a course is a credit load with a grading system of “A” >70, “B” <70 & >60, “C” <60 & >50, and “F” <50.
  - CGPA – Cumulative Grade Point Average obtained in PDE session(s) has been declared the response variable. It is split into four class values: Distinction >4.50, Credit <4.50 & >3.50, Merit < 3.50 & >2.50, and Fail <2.50.

### B. The Proposed ID3 Decision Tree Classifier

A tree in which each branch node represents a choice between some alternatives and every leaf node represents a choice is referred to as a decision tree. Decision trees are commonly used for gaining information for decision-making. The choice tree starts with a root node on which it's for users to require action. Users split each node from this node recursively, consistent with the decision tree learning algorithm. The ultimate result is a decision tree, each branch representing a possible scenario of a decision and its outcome. The three widely used decision tree learning algorithms are ID3, ASSISTANT, and C4.5.

The ID3 decision tree is a simple decision tree learning algorithm developed by Quinlan in 1986. The essential idea of the ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to check each attribute at every tree node. We introduce a metric-information gain to pick the most useful attribute for classifying a given set. To find optimal thanks to classifying a learning set, we would like to try to attenuate the questions asked (i.e., minimizing the depth of the tree). Thus, some function will be needed to measure which questions provide the foremost balance splitting. The knowledge gain metric is such a function.

### C. Measuring Impurity

Given a data table containing attributes and therefore the class of the attributes, we will measure the table's homogeneity (or heterogeneity) that supported the classes. We are saying a table is pure or homogenous if it contains only one class. If a knowledge table contains several classes, then we are saying that the table is impure or heterogeneous. There are several indices to live the degree of impurity quantitatively. The foremost well-known indices to live the degree of impurity are entropy, Gini index, and classification error. The calculation of Entropy as in (1).

$$Entropy = \sum_j^1 -P_j \log_2 P_{ij} = -P(0) * \text{Log}(P(0)) + P(1) + \log(P(1)) \quad (1)$$

A pure table's entropy (consisting of one class) is zero because the probability is 1 and  $\log(1) = 0$ . Entropy reaches a maximum value when all classes within the table have equal probability. A Gini Index equation as in (2).

$$Gini\ Index = 1 - \sum_j P_i^2 \quad (2)$$

The Gini index of a pure table containing one class is zero because the probability is 1, and  $1-1^2=0$ . Like entropy, the Gini index reaches a maximum value when all classes within the table have equal probability. The formula of classification error can be seen as in (3).

$$\text{Classification Error} = 1 - \text{Max}(P_i) \quad (3)$$

Similar to the Entropy and Gini Index, the Classification error index of a pure table (consisting of one class) is zero because the probability is 1 and  $1-\max(1)=0$ . The worth of the classification error index is usually between 0 and 1. The utmost (maximum) Gini index for a given number of classes is usually adequate to the utmost of classification error index because, for a few classes  $n$ , we set probability as equal to  $p = 1/n$ , and therefore, the maximum Gini index happens at  $1-n(1/n^2) = 1-(1/n)$ , while maximum classification error-index also happens at  $1-\max\{1/n\} = 1-(1/n)$ .

#### D. Splitting Criteria

We use the measure called Information Gain to determine the simplest attribute for a specific node within the tree. The knowledge gained expressed as Gain (S, A) of an attribute A is relative to a set of examples S – is defined as in (4).

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \quad (4)$$

Where value (A) is the set of all possible values for attribute A, and  $S_v$  is the subset of S that attribute A has value  $v$  (i.e.,  $S_v = \{s \in S \mid A(s) = v\}$ ). The first term within the equation for gain is simply the entropy of the first collection S, and therefore, the second term is that the arithmetic mean of the entropy after S is portioned using attribute A. The expected entropy is the sum of the entropies of every subset, weighted by the fraction of examples  $\frac{|S_v|}{|S|}$  that belong to Gains (S, A) is therefore the expected reduction in entropy caused by knowing the worth of attribute A. Equation (5) and (6) is use for Split information and Gain Ratio.

$$\text{Split Information}(S, A) = \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (5)$$

$$\text{Gain Ratio}(S, A) = \frac{\text{Gain}(S, A)}{\text{Split Information}(S, A)} \quad (6)$$

Choosing a replacement attribute and partitioning the training examples is now repeated for every non-terminal descendant node. Attributes incorporated higher within the tree are excluded so that any given attribute can appear at the most once along any path through the tree. This process continues for every new leaf node until either of two conditions is met. First condition is every attribute has already been included along this path through the tree [44], or the training examples related to this leaf node all have an equivalent target attribute value (i.e., their entropy is zero) [51]. The listing of the ID3 algorithm framework can be seen in Pseudocode 1.

---

#### PSEUDOCODE 1: Listing of the ID3 Algorithm Framework

---

```

ID3 (Examples, Target_Attributes, Attributes): Create a root node
if example is positive then
    return the single node tree ROOT with label = +
elseif example is negative then
    return the single-node tree Root, with label = -
end if
if a variety of predicting attributes is empty then
    return only node tree Root with label = commonest value of the target
    attribute within the examples
end if
Start:
A = The Attribute that best classifies example.
Decision Tree attribute for Root = A.
For every possible value,  $v_i$ , of A,
Add a replacement limb below Root, like the test  $A = v_i$ .
Let Example ( $v_i$ ) be the subset of example that has the worth  $v_i$  for A
If Example ( $v_i$ ) is empty; then
    below this new branch add a leaf node with label = commonest target value
    within the examples
Else below this new branch add the subtree ID3 (Example( $v_i$ ), Target_Attribute,
Attributes - {A})
end if: END

```

---



### III. Result and Discussion

A dataset of 50 students was used in this study (Table 2) as obtained/retrieved from the Professional Diploma in Education, Federal College of Education (Technical) Asaba, PDE-Technical Option from session 2017 to 2020 [74].

Table 2. Dataset from PDE Technical from 2016 - 2019

S.No	CGP	TP	ASS	GA	ATT	EP	CGPA
1	Distinction	Distinction	Yes	Yes	Good	Yes	Distinction
2	Distinction	Credit	Yes	No	Good	Yes	Distinction
3	Distinction	Credit	No	No	Average	No	Distinction
4	Credit	Distinction	No	No	Good	Yes	Distinction
5	Credit	Credit	No	Yes	Good	Yes	Distinction
6	Merit	Credit	No	No	Average	Yes	Distinction
7	Merit	Credit	No	No	Poor	Yes	Credit
8	Credit	Merit	Yes	Yes	Average	No	Distinction
9	Merit	Merit	No	No	Poor	No	Merit
10	Credit	Credit	Yes	Yes	Good	No	Distinction
11	Distinction	Distinction	Yes	Yes	Good	Yes	Distinction
12	Distinction	Credit	Yes	Yes	Good	Yes	Distinction
13	Distinction	Credit	Yes	No	Good	No	Distinction
14	Credit	Distinction	Yes	Yes	Good	No	Distinction
15	Distinction	Credit	Yes	Yes	Average	Yes	Distinction
16	Distinction	Credit	Yes	Yes	Poor	Yes	Credit
17	Credit	Credit	Yes	Yes	Good	Yes	Credit
18	Credit	Credit	Yes	Yes	Poor	Yes	Credit
19	Merit	Credit	No	Yes	Good	Yes	Credit
20	Credit	Merit	Yes	No	Average	Yes	Credit
21	Merit	Credit	No	Yes	Poor	No	Merit
22	Merit	Merit	Yes	Yes	Average	Yes	Merit
23	Merit	Merit	No	No	Average	Yes	Merit
24	Merit	Merit	Yes	Yes	Good	Yes	Credit
25	Merit	Merit	Yes	Yes	Poor	Yes	Merit
26	Merit	Merit	No	No	Poor	Yes	Fail
27	Distinction	Distinction	Yes	Yes	Good	Yes	Distinction
28	Credit	Distinction	Yes	Yes	Good	Yes	Credit
29	Distinction	Credit	Yes	Yes	Good	Yes	Credit
30	Distinction	Distinction	Yes	Yes	Average	Yes	Credit
31	Distinction	Distinction	No	No	Good	Yes	Credit
32	Credit	Credit	Yes	Yes	Good	Yes	Credit
33	Credit	Credit	No	Yes	Average	Yes	Merit
34	Credit	Distinction	No	No	Good	Yes	Merit
35	Distinction	Credit	No	Yes	Average	Yes	Merit
36	Credit	Merit	No	No	Average	Yes	Merit
37	Merit	Credit	Yes	No	Average	Yes	Merit
38	Merit	Credit	No	Yes	Poor	Yes	Fail
39	Credit	Credit	No	Yes	Poor	Yes	Merit
40	Merit	Merit	No	No	Good	No	Merit
41	Merit	Merit	No	Yes	Poor	Yes	Fail
42	Merit	Merit	No	No	Poor	No	Fail
43	Distinction	Distinction	Yes	Yes	Good	Yes	Credit
44	Distinction	Distinction	Yes	Yes	Average	Yes	Credit
45	Credit	Distinction	Yes	Yes	Average	Yes	Merit
46	Merit	Merit	Yes	Yes	Average	No	Fail
47	Distinction	Merit	No	Yes	Poor	Yes	Fail
48	Merit	Merit	No	No	Poor	Yes	Fail
49	Credit	Credit	Yes	Yes	Good	Yes	Credit
50	Merit	Distinction	No	No	Poor	No	Fail

To compute the knowledge gain for A relative to S – we compute the Entropy of S. But, S is the set of all of the 50-examples with value 12=“Distinction”, 15=“Credit”, 17=“Merit”, and 6=“Fail”. So, we have:

$$\begin{aligned}
 Entropy(S) &= -D_{instinction} \log(D_{instinction}) - C_{redit} \log(C_{redit}) - M_{erit} \log(M_{erit}) - F_{ail} \log(F_{ail}) \\
 &= -\frac{12}{50} \log \left[ \frac{12}{50} \right] - \frac{15}{50} \log \left[ \frac{15}{50} \right] - \frac{17}{50} \log \left[ \frac{17}{50} \right] - \frac{6}{50} \log \left[ \frac{6}{50} \right] = 1.911
 \end{aligned}$$

We use the measure called Information Gain to determine the simplest attribute for a specific node within the tree. The knowledge gain, Gain (S, A) of an attribute A, relative to a set of examples S:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \left[ \frac{|S_{\text{first}}|}{|S|} \text{Entropy}|S_{\text{first}}| \right] - \left[ \frac{|S_{\text{second}}|}{|S|} \text{Entropy}|S_{\text{second}}| \right] - \left[ \frac{|S_{\text{third}}|}{|S|} \text{Entropy}|S_{\text{third}}| \right] - \left[ \frac{|S_{\text{fail}}|}{|S|} \text{Entropy}|S_{\text{fail}}| \right]$$

CGP has the highest gain. Thus, it is used as the root node as shown in Figure 2. Table 3 represents the Gain Values. Table 4 shows the Split information values, while Table 5 represents the Gain Ratio.

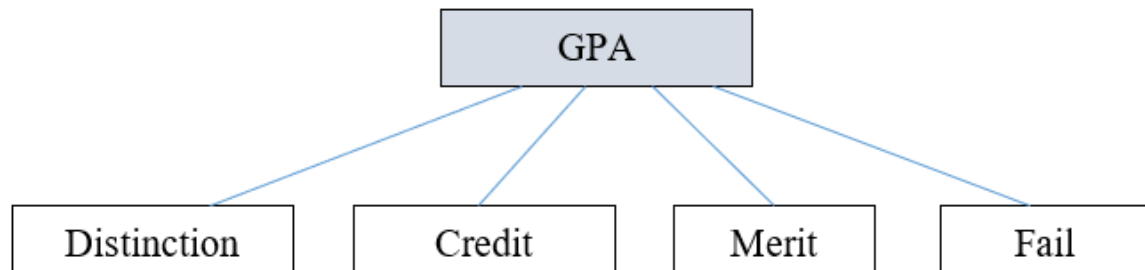


Fig. 2. The steps of extracting knowledge from data

Table 3. Gain values

Split information	Value
Split (S, CGP)	1.448442
Split (S, TP)	1.597734
Split (S, ASS)	1.744987
Split (S, GA)	1.91968
Split (S, ATT)	1.511673
Split (S, EP)	1.510102

Table 4. Split information

Gain	Value
Gain(S, CGP)	1.690616
Gain(S, TP)	1.602740
Gain(S, ASS)	0.995378
Gain(S, GA)	0.924819
Gain(S, ATT)	1.560956
Gain(S, EP)	0.826746

Table 5. Gain ratio

Gain Ratio	Value
Gain Ratio (S, CGP)	0.355674
Gain Ratio (S, TP)	0.229
Gain Ratio (S, ASS)	0.125289
Gain Ratio (S, GA)	0.022887
Gain Ratio (S, ATT)	0.298968
Gain Ratio (S, EP)	0.30032

This process continues until all data is classified ideally or out of attributes. The knowledge represented by the decision tree can be extracted and represented in the form of IF-THEN rules as denoted in the Pseudocode 2.

One classification rule can be generated for each path from each terminal node to the root node. The pruning technique was executed by removing nodes with less than the desired number of objects. IF-THEN rules may be easier to understand.



**PSEUDOCODE 2: Listing of If-Then rules for the Decision Tree**


---

```

if CGP = "Distinction" AND ATT = "Good" AND TP = "A" or "Credit" then
  ASS = "Yes"
else-if
if CGP = "Distinction" AND TP = "A" AND ATT = "Good" OR "Average" then
  ASS = "Yes"
else-if
if CGP = "Credit" AND ATT = "Good" AND ASS = "Yes" then
  CGPA = "Distinction"
else-if
if CGP = "Credit" AND TP = "B" AND EP = "A" then
  CGPA = "Credit"
else-if
if IF CGP = "Merit" AND TP = "A" OR "B" AND ATT = "Good" OR "Average" then
  ASS = "Yes"
else-if
if CGP = "Merit" AND ASS = "No" AND ATT = "Average" then
  CGPA = "Merit"
else-if
if CGP = "Fail" AND TP = "F" AND ATT = "Poor" then
  CGPA = "Fail"
else-if

```

---

**IV. Conclusion**

The classification task is used on the scholar database to predict the student's division supported by the previous database. As many approaches as can be used for data classification, the decision tree method is used for this study. Also, data like the Cumulative Grade Point, Teaching Practice marks, Assignments, General Aptitude, Attendance, Education Project marks, and Cumulative Grade Point Average – were collected from the student's previous database to predict the performance at the top of the semester. This study will help the scholars and teachers enhance the division of the scholar. This study also will work to spot students who need special attention to reduce the fail ratio and take appropriate action for subsequent semester's examinations.

**Declarations***Author contribution*

All authors contributed equally as the main contributor of this paper. All authors read and approved the final paper.

*Funding statement*

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

*Conflict of interest*

The authors declare no known conflict of financial interest or personal relationships that could have appeared to influence the work reported in this paper.

*Additional information*

Reprints and permission information are available at <http://journal2.um.ac.id/index.php/keds>.

Publisher's Note: Department of Electrical Engineering and Informatics - Universitas Negeri Malang remains neutral with regard to jurisdictional claims and institutional affiliations.

**References**

- [1] A. Ifeka and A. Akinbobola, "Trend Analysis of Precipitation in Some Selected Stations in Anambra State," *Atmos. Clim. Sci.*, vol. 05, no. 01, pp. 1–12, 2015.
- [2] M. I. Akazue, R. E. Yoro, B. O. Malasowe, O. Nwankwo, and A. A. Ojugo, "Improved services traceability and management of a food value chain using block-chain network: a case of Nigeria," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 29, no. 3, pp. 1623–1633, 2023.

- [3] A. A. Ojugo, P. O. Ejeh, C. C. Odiakoese, A. O. Eboka, and F. U. Emordi, "Improved distribution and food safety for beef processing and management using a blockchain-tracer support framework," *Int. J. Informatics Commun. Technol.*, vol. 12, no. 3, p. 205, Dec. 2023.
- [4] R. E. Yoro, F. O. Aghware, B. O. Malasowe, O. Nwankwo, and A. A. Ojugo, "Assessing contributor features to phishing susceptibility amongst students of petroleum resources varsity in Nigeria," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 2, p. 1922, Apr. 2023.
- [5] R. E. Yoro, F. O. Aghware, M. I. Akazue, A. E. Ibor, and A. A. Ojugo, "Evidence of personality traits on phishing attack menace among selected university undergraduates in Nigerian," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 2, p. 1943, Apr. 2023.
- [6] S. Drummond, K. Sudduth, A. Joshi, S. Birrell, and S. Kitchen, "Statistics and neural method for site specific yield prediction," *Trans. ASAE*, vol. 46, no. 1, pp. 23–32, 2003.
- [7] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemom. Intell. Lab. Syst.*, vol. 83, no. 2, pp. 83–90, Sep. 2006.
- [8] A. A. Ojugo and A. O. Eboka, "Memetic algorithm for short messaging service spam filter using text normalization and semantic approach," *Int. J. Informatics Commun. Technol.*, vol. 9, no. 1, p. 9, 2020.
- [9] Q. Li et al., "An Enhanced Grey Wolf Optimization Based Feature Selection Wrapped Kernel Extreme Learning Machine for Medical Diagnosis," *Comput. Math. Methods Med.*, vol. 2017, pp. 1–15, 2017.
- [10] J. W. Hatfield, C. R. Plott, and T. Tanaka, "Understanding Price Controls and Nonprice Competition with Matching Theory," *Am. Econ. Rev.*, vol. 102, no. 3, pp. 371–375, May 2012.
- [11] A. A. Ojugo and R. E. Yoro, "Migration Pattern As Threshold Parameter In The Propagation of The Covid-19 Epidemic Using An Actor-Based Model for SI-Social Graph," *JINAV J. Inf. Vis.*, vol. 2, no. 2, pp. 93–105, Mar. 2021.
- [12] A. A. Ojugo and O. Nwankwo, "Spectral-Cluster Solution For Credit-Card Fraud Detection Using A Genetic Algorithm Trained Modular Deep Learning Neural Network," *JINAV J. Inf. Vis.*, vol. 2, no. 1, pp. 15–24, Jan. 2021.
- [13] M. I. Akazue, A. A. Ojugo, R. E. Yoro, B. O. Malasowe, and O. Nwankwo, "Empirical evidence of phishing menace among undergraduate smartphone users in selected universities in Nigeria," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 3, pp. 1756–1765, Dec. 2022.
- [14] S. Carbó, J. F. De Guevara, D. Humphrey, and J. Maudos, "Estimating the intensity of price and non-price competition in banking," *Banks Bank Syst.*, vol. 4, no. 2, pp. 4–19, 2009.
- [15] L. A. Belanche and F. F. González, "Review and Evaluation of Feature Selection Algorithms in Synthetic Problems," *Inf. Fusion*, vol. 23, pp. 34–54, Jan. 2011.
- [16] Z. Karimi, M. Mansour Riahi Kashani, and A. Harounabadi, "Feature Ranking in Intrusion Detection Dataset using Combination of Filtering Methods," *Int. J. Comput. Appl.*, vol. 78, no. 4, pp. 21–27, Sep. 2013.
- [17] A. Karim, S. Azam, B. Shanmugam, K. Kannoopatti, and M. Alazab, "A Comprehensive Survey for Intelligent Spam Email Detection," *IEEE Access*, vol. 7, pp. 168261–168295, 2019.
- [18] N. Tomar and A. K. Manjhar, "A Survey on Data Mining Optimization Techniques," *IJSTE-International J. Sci. Technol. Eng. J.*, vol. 2, no. 06, pp. 130–133, 2015.
- [19] A. Goldstein, L. Fink, A. Meitin, S. Bohadana, O. Lutenberg, and G. Ravid, "Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge," *Precis. Agric.*, vol. 19, no. 3, pp. 421–444, Jun. 2018.
- [20] A. A. Ojugo and D. A. Oyemade, "Boyer moore string-match framework for a hybrid short message service spam filtering technique," *IAES Int. J. Artif. Intell.*, vol. 10, no. 3, pp. 519–527, 2021.
- [21] U. Usman, "Effects of Price & Non-Price Competition of Consumers Effects of Pricing and Non-Pricing Competition on Consumer Submitted By: Umair Usman Ghani Submitted To: Sir Raja Rub Nawaz Dated Preston University - Karachi Main Campus," pp. 1–16, 2014.
- [22] F. Shirbani and H. Soltanian Zadeh, "Fast SFFS-Based Algorithm for Feature Selection in Biomedical Datasets," *Amirkabir Int. J. Sci. Res.*, vol. 45, no. 2, pp. 43–56, 2013.
- [23] A. A. Ojugo, A. O. Eboka, R. E. Yoro, M. O. Yerokun, and F. N. Efozia, "Hybrid model for early diabetes diagnosis," *Math. Comput. Ind.*, vol. 50, no. 3–5, pp. 55–65, 2015.
- [24] G. B. Dela Cruz, B. D. Gerardo, and B. T. Tanguilig III, "Agricultural Crops Classification Models Based on PCA-GA Implementation in Data Mining," *Int. J. Model. Optim.*, vol. 4, no. 5, pp. 375–382, Oct. 2014.
- [25] Y. Shiokawa, T. Misawa, Y. Date, and J. Kikuchi, "Application of Market Basket Analysis for the Visualization of Transaction Data Based on Human Lifestyle and Spectroscopic Measurements," *Anal. Chem.*, vol. 88, no. 5, pp. 2714–2719, 2016.
- [26] A. Patil and P. Gupta, "A review on up-growth algorithm using association rule mining," in *2017 International Conference on Computing Methodologies and Communication (ICCMC)*, Jul. 2017, pp. 96–99.
- [27] H. W. Ahmad, S. Zilles, H. J. Hamilton, and R. Dosselmann, "Prediction of retail prices of products using local competitors," *Int. J. Bus. Intell. Data Min.*, vol. 11, no. 1, pp. 19–30, 2016.
- [28] M. Brindlmayer, R. Khadduri, A. Osborne, A. Briansó, and E. Cupito, "Prioritizing learning during covid-19: The Most Effective Ways to Keep Children Learning During and Post-Pandemic," *Glob. Educ. Evid. Advis. Panel*, no. January, pp. 1–21, 2022.
- [29] V.-D. Nguyen, D.-N. Tran, H.-H. Tran, T.-N. Phan, T. Danh, and H.-N. Tran, "Blended Learning Model-Based Local Education for Vietnamese Primary School Students," *Rev. Int. Geogr. Educ.*, vol. 11, no. 8, pp. 1684–1694, 2022.
- [30] D. Nilam, W. Sari, and M. Mulu, "Explorative study on the application of learning model in virtual classroom during Covid-19 pandemic at the school of Yogyakarta Province," *Proceeding Int. Webinar Educ. 2020 Umsurabaya*, pp. 54–64, 2020.
- [31] D. L. Chen, S. Ertac, T. Evgeniou, X. Miao, A. Nadaf, and E. Yilmaz, "Grit and Academic Resilience During the Covid-19 Pandemic," *SSRN Electron. J.*, 2022.
- [32] E. Haiping, N. Kadhila, and L. M. Josua, "Using Digital Technology in Transforming Assessment in Higher Education Institutions beyond COVID-19," *Creat. Educ.*, vol. 13, no. 07, pp. 2157–2167, 2022.
- [33] H. Patrinos, E. Vegas, and R. Carter-Rau, "An Analysis of COVID-19 Student Learning Loss," *Educ. Glob. Pract. Policy Res. Work. Pap.* 10033, vol. 10033, no. May, pp. 1–31, 2022.
- [34] F. Agostinelli, M. Doepke, G. Sorrenti, and F. Zilibotti, "When the great equalizer shuts down: Schools, peers, and parents in pandemic times," *J. Public Econ.*, vol. 206, p. 104574, Feb. 2022.
- [35] U. Christian and M. Author, "The Influence of Covid-19 on Good Governance and Democratic Behavior in Nigeria," *Int. J. Arts Soc. Sci.*, vol. 5, no. July, pp. 50–57, 2022.

- [36] I. M. Ugochukwu-Ibe and E. Ibeke, "E-learning and Covid-19 - The Nigerian experience: Challenges of teaching technical courses in tertiary institutions," *CEUR Workshop Proc.*, vol. 2872, no. May, pp. 46–51, 2021.
- [37] W. C. Kolberg, "Marketing Mix Theory: Integrating Price and Non-Price Marketing Strategies," *SSRN Electron. J.*, no. 1993, pp. 1–35, 2011.
- [38] A. A. Ojugo and O. Nwankwo, "Tree-classification Algorithm to Ease User Detection of Predatory Hijacked Journals: Empirical Analysis of Journal Metrics Rankings," *Int. J. Eng. Manuf.*, vol. 11, no. 4, pp. 1–9, Aug. 2021.
- [39] A. E. Ibor, E. B. Edim, and A. A. Ojugo, "Secure Health Information System with Blockchain Technology," *J. Niger. Soc. Phys. Sci.*, vol. 5, no. 992, pp. 1–8, 2023.
- [40] F. O. Aghware, R. E. Yoro, P. O. Ejeh, C. Odiakoese, F. U. Emordi, and A. A. Ojugo, "Sentiment Analysis in Detecting Sophistication and Degradation Cues in Malicious Web Contents," *Kongzhi yu Juece/Control Decis.*, vol. 38, no. 01, pp. 653–665, 2023.
- [41] K. Vassil, M. Solvak, P. Vinkel, A. H. Trechsel, and R. M. Alvarez, "The diffusion of internet voting. Usage patterns of internet voting in Estonia between 2005 and 2015," *Gov. Inf. Q.*, vol. 33, no. 3, pp. 453–459, Jul. 2016.
- [42] W. Pieters, "Acceptance of Voting Technology: Between Confidence and Trust," in *International Conference on Trust Management*, 2006, pp. 283–297.
- [43] S. Okuyama, S. Tsuruoka, H. Kawanaka, and H. Takase, "Interactive Learning Support User Interface for Lecture Scenes Indexed with Extracted Keyword from Blackboard," *Aust. J. Basic Appl. Sci.*, vol. 8, no. 4, pp. 319–324, 2014.
- [44] S. Chouhan, D. Singh, and A. Singh, "An Improved Feature Selection and Classification using Decision Tree for Crop Datasets," *Int. J. Comput. Appl.*, vol. 142, no. 13, pp. 5–8, May 2016.
- [45] J. Obasi, Nwele, N. Amuche N, and U. Elias A., "Economics of Optimizing Value Chain in Agriculture Sector of Nigeria through Mechanised Crop Processing and Marketing," *Asian J. Basic Sci. Res.*, vol. 02, no. 01, pp. 80–92, 2020.
- [46] D. Acemoglu, K. Bimpikis, and A. Ozdaglar, "Price and capacity competition: Extended abstract," *44th Annu. Allert. Conf. Commun. Control. Comput.* 2006, vol. 3, no. December, pp. 1307–1309, 2006.
- [47] E. Oyeboode, K. Adekalu, and S. Akinboro, "Development of rainfall-runoff forecast model," *J. Res. Natl. Dev.*, vol. 8, no. 2, pp. 56–66, 2011.
- [48] A. A. Ojugo, C. O. Obruché, and A. O. Eboka, "Quest For Convergence Solution Using Hybrid Genetic Algorithm Trained Neural Network Model For Metamorphic Malware Detection," *ARRUS J. Eng. Technol.*, vol. 2, no. 1, pp. 12–23, Nov. 2021.
- [49] A. A. Ojugo, C. O. Obruché, and A. O. Eboka, "Empirical Evaluation for Intelligent Predictive Models in Prediction of Potential Cancer Problematic Cases In Nigeria," *ARRUS J. Math. Appl. Sci.*, vol. 1, no. 2, pp. 110–120, Nov. 2021.
- [50] G. G. Akin, A. F. Aysan, G. I. Kara, and L. Yildiran, "The failure of price competition in the Turkish credit card market," *Emerg. Mark. Financ. Trade*, vol. 46, no. SUPPL. 1, pp. 23–35, 2010.
- [51] D. O. Oyewola, E. G. Dada, N. J. Ngozi, A. U. Terang, and S. A. Akinwumi, "COVID-19 Risk Factors, Economic Factors, and Epidemiological Factors nexus on Economic Impact: Machine Learning and Structural Equation Modelling Approaches," *J. Niger. Soc. Phys. Sci.*, vol. 3, no. 4, pp. 395–405, 2021.
- [52] J. H. Jeong et al., "Random Forests for Global and Regional Crop Yield Predictions," *PLoS One*, vol. 11, no. 6, p. e0156571, Jun. 2016.
- [53] X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques," *Comput. Electron. Agric.*, vol. 121, pp. 57–65, Feb. 2016.
- [54] A. A. Ojugo and O. D. Otakore, "Intelligent cluster connectionist recommender system using implicit graph friendship algorithm for social networks," *IAES Int. J. Artif. Intell.*, vol. 9, no. 3, p. 497–506, 2020.
- [55] T. Avinadav, "The effect of decision rights allocation on a supply chain of perishable products under a revenue-sharing contract," *Int. J. Prod. Econ.*, vol. 225, p. 107587, Jul. 2020.
- [56] F. O. Aghware, R. E. Yoro, P. O. Ejeh, C. C. Odiakoese, F. U. Emordi, and A. A. Ojugo, "DeLClustE: Protecting Users from Credit-Card Fraud Transaction via the Deep-Learning Cluster Ensemble," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, pp. 94–100, 2023.
- [57] M. Armstrong and J. Vickers, "Patterns of Price Competition and the Structure of Consumer Choice," *MPRA Pap.*, vol. 1, no. 98346, pp. 1–40, 2020.
- [58] K. Parsons, A. McCormac, M. Pattinson, M. Butavicius, and C. Jerram, "The design of phishing studies: Challenges for researchers," *Comput. Secur.*, vol. 52, pp. 194–206, Jul. 2015.
- [59] S. Girish Patil, P. Shahaji, N. Niles, G. Kishore, and R. Gupta, "Traceability Based Value Chain Management in Meat Sector for Achieving Food Safety and Augmenting Exports, 2022.
- [60] C. Li, N. Ding, H. Dong, and Y. Zhai, "Application of Credit Card Fraud Detection Based on CS-SVM," *Int. J. Mach. Learn. Comput.*, vol. 11, no. 1, pp. 34–39, 2021.
- [61] V. Umarani, A. Julian, and J. Deepa, "Sentiment Analysis using various Machine Learning and Deep Learning Techniques," *J. Niger. Soc. Phys. Sci.*, vol. 3, no. 4, pp. 385–394, 2021.
- [62] B. O. Malasowe, M. I. Akazue, E. A. Okpako, F. O. Aghware, A. A. Ojugo, and D. V. Ojie, "Adaptive Learner-CBT with Secured Fault-Tolerant and Resumption Capability for Nigerian Universities," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 135–142, 2023.
- [63] S. Khaki, L. Wang, and S. V. Archontoulis, "A CNN-RNN Framework for Crop Yield Prediction," *Front. Plant Sci.*, vol. 10, no. January, pp. 1–14, 2020.
- [64] S. Khaki and L. Wang, "Crop Yield Prediction Using Deep Neural Networks," *Front. Plant Sci.*, vol. 10, May 2019.
- [65] A. D. Bhavani and N. Mangla, "A Novel Network Intrusion Detection System Based on Semi-Supervised Approach for IoT," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 4, pp. 207–216, 2023.
- [66] M. Sharma, "A Survey of Email Spam Filtering Methods," *Int. Conf. "New Trends Stat. Optim."*, vol. 7, no. 6, pp. 14–21, 2018.
- [67] Z. Sun, S. Sun, J. Zhao, B. Ai, and Q. Yang, "Detection of Massive Oil Spills in Sun Glint Optical Imagery through Super-Pixel Segmentation," *J. Mar. Sci. Eng.*, vol. 10, no. 11, p. 1630, 2022.
- [68] S. Do, K. D. Song, and J. W. Chung, "Basics of Deep Learning: A Radiologist's Guide to Understanding Published Radiology Articles on Deep Learning," *Korean J. Radiol.*, vol. 21, no. 1, pp. 33–41, 2020.
- [69] A. S. Pillai, "Multi-Label Chest X-Ray Classification via Deep Learning," *J. Intell. Learn. Syst. Appl.*, vol. 14, pp. 43–56, 2022.
- [70] S. K. Datta, M. A. Shaikh, S. N. Srihari, and M. Gao, "Soft-Attention Improves Skin Cancer Classification Performance," May 2021.

- [71] Y. Kang, M. Ozdogan, X. Zhu, Z. Ye, C. Hain, and M. Anderson, “Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest,” *Environ. Res. Lett.*, vol. 15, no. 6, p. 064005, Jun. 2020.
- [72] A. A. Ojugo and R. E. Yoro, “Extending the three-tier constructivist learning model for alternative delivery: ahead the COVID-19 pandemic in Nigeria,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 21, no. 3, p. 1673, Mar. 2021.
- [73] A. A. Ojugo and R. E. Yoro, “Forging a deep learning neural network intrusion detection framework to curb the distributed denial of service attack,” *Int. J. Electr. Comput. Eng.*, vol. 11, no. 2, pp. 1498–1509, 2021.
- [74] A. A. Ojugo, M. I. Akazue, P. O. Ejeh, C. Odiakaose, and F. U. Emordi, “DeGATraMoNN : Deep Learning Memetic Ensemble to Detect Spam Threats via a Content-Based Processing,” *Kongzhi yu Juece/Control Decis.*, vol. 38, no. 01, pp. 667–678, 2023.