International Journal of Scientific Research in Computer Science and Engineering
Vol.**12**, Issue.**4**, pp.**54-65**, August **2024**
E-ISSN: 2320-7639
Available online at: www.isroset.org

Research Article

Association Rule Mining on Likert's Scale Data using a Novel Attributes Pruning Technique

Ogedengbe M.T.[1*], Junaidu S.B.[2], Kana A.F.D.[3]

[1]Dept. of Computer Science, Joseph Sarwuan Tarka University, Makurdi, Nigeria
[2,3]Dept. of Computer Science, Ahmadu Bello University, Zaria, Nigeria

**Corresponding Author: matt.oged@uam.edu.ng*

Received: 22/Jun/2024; **Accepted**: 24/Jul/2024; **Published**: 31/Aug/2024

Abstract— In recent decades, the Apriori algorithm has emerged as a powerful tool for generating meaningful insights and supporting effective decision-making in data science. Traditionally a binary mining tool, Apriori is highly efficient in analysing transaction datasets in market basket analysis to uncover customer purchasing patterns. When applied to Likert scale datasets, however, it requires discretizing item attributes into binary values, which can result in significant information loss. This study proposes a novel mining technique called Common Question Attribute Pruning (CQAP), which enhances the standard Apriori algorithm by extending its capabilities to process 5-point Likert scale datasets without the need for attribute discretization, thereby preserving the ordinal nature of the respondents' opinions. The key innovation of this technique lies in its ability to represent all five points of the Likert scale within the Apriori framework, without converting them to Boolean transaction data (0s and 1s). The modified Apriori algorithm, termed the Extended Apriori Algorithm (Ext-AA), generates candidate sets by treating question-value (q,v) pairs as data points. During the candidate joining phase, any combinations with common question attributes are pruned. This approach introduces a new perspective on defining support count, minimum support threshold metric, and minimum confidence threshold metric, which helps in filtering out infrequent candidate sets and the determination of the strength of the association rules derived from the sample dataset. In experimental evaluations on sample datasets, the Ext-AA produced 8 strong rules, whereas the standard Apriori algorithm generated 135 rules which is 89.63% rule reduction after pruning. These results demonstrate the superior performance potential of CQAP technique against the state of art Apriori algorithm on the evaluated sample datasets.

Keywords— Apriori; Pruning; Likert scale; Questionnaire; Candidate-set; Itemset; Support; Confidence.

1. Introduction

Data mining involves the extraction and analysis of large datasets from various application areas to uncover useful patterns and relationships, aiding in effective decision-making [1]. In today's era of big data and information technology, vast amounts of data are generated by business organizations, government activities, financial institutions, healthcare systems, and more. Processing such extensive data manually within a limited time frame is impractical for efficient decision-making [2][3]. Therefore, it is essential to employ efficient data mining techniques to analyze, classify, and summarize these datasets, to reveal the hidden patterns and relationships that facilitate meaningful decisions [4]. The primary challenge lies in selecting appropriate mining tools for specific datasets. Utilizing the right set of tools for data mining should yield accurate results while maintaining the integrity of the mining process [5][6]. Discovering relationships among data in large databases, such as sales transactions, is crucial for organizations because the associations and connections among data values which provides significant business value [7]. Apriori algorithm is the conventional mining tool for association rules discovery [8][9]. It is categorized as unsupervised learning technique proposed by Agrawal in 1993 for market basket analysis, specifically for observing customer purchasing patterns and behavior mostly in market basket analysis [10]. Over time, AR has been widely adopted in various research areas including image processing, customer relationship management, aviation mining [11][12][42], text mining, information visualization, educational mining, and air pollution mining [13]. The Apriori extracts meaningful information from data warehouses or databases by identifying items which are frequent and generating candidate itemsets to produce association rules using minimum support and confidence measure [14][15]. The effectiveness of these measures significantly impacts the generation of association rules [16]. Strong rule is that rule whose confidence value is at least the value of the minimum confidence [17]. Association rule mining (ARM) focus on unveiling the associations or correlations among items in a large transactional database [42]. This process can investigate various attitudes and perceptions by exploring factors responsible for changes in human perception. Attitudes can be measured using the Likert scale, a psychometric tool used for assessing behavioral perceptions or ideas within a given research domain. The Likert scale, a prevalent tool for surveying human attitudes, is used to gather responses on specific questions or related problems such as societal service levels, healthcare services, and customer satisfaction [18]. Developed by Rensis Likert, this scale quantifies respondents' attitudes by rating their level of agreement with given statements [19][20][21]. It enables researchers to investigate both qualitatively and quantitatively the opinions or perceptions of respondents. Typically, respondents have ranges of choice to be selected to express their agreement or disagreement level on a 5-points or 7-points scale, which ranges from “strongly agree” to “strongly disagree”, depending on the researchers' design [22]. Despite its widespread use and advantages, the Likert scale has analytical limitations, primarily due to its qualitative nature [23][24]. To enhance analytical possibilities, it is often combined with other data mining tools [25]. However, this paper presents an enhanced Apriori algorithm designed to mine Likert scale data without the need for discretizing its values into Boolean form. Traditional discretization can result in a loss of detail in respondents' perceptions, as it reduces their nuanced opinions to binary values [26]. Our proposed approach maintains the integrity of the original responses, preserving the full spectrum of opinions from 'strongly disagree' to 'strongly agree' [27] This new technique, which we call Common Question Attribute Pruning (CQAP), extends the Apriori algorithm to handle the full range of Likert scale responses, ensuring that the richness of the data is retained and more meaningful association rules are generated.

The remaining sections of this article are presented as follows, Section 1 comprises the introduction of association rule mining as used in data mining, Section 2 focused on the review of the related work, while Section 3 discussed the research methodology and the pseudo-codes of the proposed Ex-AA. Section 4 presents results obtained from the association rules generated, and finally, Section 5 presents the concluding part of the research work with possible future directions.

2. Related Work

Data mining is a technique used to extract meaningful knowledge from extensive database [41]. The technique involves searching, analysing, and evaluating large datasets to uncover new patterns and relationships, ultimately aiding in knowledge acquisition and informed decision making. The knowledge derived from data mining should meet three key criteria: accuracy, comprehensibility, and interestingness. By identifying the most relevant patterns, data mining extracts meaningful ideas/information from voluminous data stored in web repository, databases, data warehouses, web repositories, and dynamic data streams [28]. The primary objective of data mining involves transforming voluminous data to actionable knowledge. This process involves several steps, including data preprocessing, data cleaning, and pattern discovery using various algorithms and techniques. The discovered patterns can then be used to support meaningful and decisive analysis in various research domains such as finance, business, healthcare, and many more [29]. Data mining serves as a critical tool in the era of big data, enabling organizations to leverage their data for strategic advantage. The effective use of data mining can lead to improved operational efficiency, better customer insights, and enhanced overall performance [30].

Apriori algorithm was applied to mine students' scores from an entrance examination in the English section, using Gardner’s attitude test battery questionnaire model. This study employed a 6-point Likert scale questionnaire with 25 multiple-choice questions and included 520 students in the survey. The questionnaire attributes were first discretised and then converted to binary values (0-1). The results revealed 15 significant association rules based on the given minimum confidence value. These studies illustrate the versatility of the Apriori algorithm in different contexts, from predicting cyber scam vulnerability to analysing educational performance, while also highlighting the challenges of discretising Likert scale data for effective rule mining [31].

Association analysis was conducted on Masters in Computer Application (MCA) students to determine the courses required for the MCA program in [32]. The dataset used comprised 57 instances with two types of attributes: conditional attributes representing courses offered, and decision attributes representing the class of degree. The conditional attributes were discretised into nine grades, while the decision attributes were discretised into six grades, both over a range of 0-100 marks. The mining algorithm did not specify the minimum support but instead specified the number of desired rules. Results showed that 18 interesting rules were generated, highlighting the importance of a sound knowledge of programming courses for students enrolling in the MCA program.

In prior research [33], an enhanced version of the Apriori-Gen algorithm was employed to analyse a questionnaire consisting of 47 questions related to environmental, learner, and teacher factors. Preceding association analysis, the questionnaire data was discretized into transactional data. The algorithm was utilized to assess the efficacy of conceptual and political courses within universities by correlating questionnaire responses. Parameters such as support value and confidence coefficient were set as 0.2 and 0.9 respectively, to derive relevant rules for correlating teacher, student, and environmental factors. The analysis yielded approximately 160,000 rules, considered unwieldy for practical use, prompting a reduction to 75 effective rules. These rules were identified as instrumental in enhancing university courses.

A combined row-wise generation technique was used in [1] to eliminate unnecessary itemsets by some researchers. Their technique employed all frequent itemsets in each transaction row to generate possible combinations of items. A minimum support threshold (MST) was adopted to find minimum support without user specification, the threshold was calculated as the average of the minimum and maximum frequent itemsets in the transaction database $MST = \frac{1}{2}(min + max)$. The results revealed that the association rules generated were based on the most frequent itemsets in the dataset.

A romance scam prediction model was proposed in [34] using Routine Activity Theory (RAT) and Apriori algorithm to identify the major factors contributing to cyber romance scams in Malaysia. The study utilized Likert scale data, with attributes such as marital status, age, education level, monthly income, computer skills, and cyber-fraud awareness. The experiment was experimented on Waikato Environment for Knowledge Analysis (WEKA) tool, 0.1 was specified as support threshold value and 1.0 was specified as the confidence threshold value for extracting association rules. Results indicated that individuals aged 25-45 years were particularly vulnerable to cyber romance crimes. Additionally, the result also revealed that, those with non-income and lacking computer skills were more likely to fall victim to these scams.

In another study [35], student performance was predicted using the Apriori algorithm to uncover hidden patterns within academic data. The investigation focused on 15 instances and 5 attributes, including attendance, lab work, class test grades, assignments, and last semester grades. Attributes were formatted accordingly for mining purposes, and association analyses revealed nuanced insights into student performance. Notably, the study found that attendance had the least impact on student performance compared to other factors, with only 13% of students with good attendance achieving an A grade.
Association rule mining was developed in [36] for analysing the financial mathematics subject with the objective of identifying various factors affecting students' performance. The dataset contained 273 instances presented with a 5-point Likert scale questionnaire to capture respondents' opinions. The students' responses were generated across 18 different courses, classified as independent (other mathematical courses) variables against the dependent (financial mathematics course) variables. The 5-scaled attributes were converted to binary attributes using Apriori algorithm, with each attribute set as a single factor for the mining process. The minimum support threshold was set to 0.02 to reveal rare rules. This resulted in 62 rules, showing associations between the attributes and the dependent variables. The accuracy and strength of each relationship were determined based on the predefined minimum confidence and lift values.

In [37], spatial heterogeneity was considered as an aspect of environmental heterogeneity in participants' visual landscape preferences. The approach aimed to find consensus on perceptual parameters of different landscape character types influencing participants' visual preferences using a 5-point Likert scale questionnaire. For determining visual landscape preferences, 243 photographs were scored with a 5-point Likert scale by five landscape architects. Frequent itemsets and candidate sets were generated based on the minimum support constraint. The study found that landscapes such as lakes, forests, natural vegetation, wildlife, road landscapes, and historical structures deserved more protection and had higher visual preferences based on the association rules generated by the minimum confidence constraint.

While previous research has effectively utilized the Apriori algorithm on Likert scale datasets to generate association rules, necessitating a preprocessing step of discretization, however, our study proposes a novel approach. We introduce a methodology to directly mine Likert scale datasets without the need for discretization, thereby streamlining the analysis process [22].

3. Theory/Calculation

In ordinal Likert mining, respondents' perspectives were assessed through the questionnaires, and are quantified using 5-scaled points, from "Strongly agree" (5) to "Strongly Disagree" (1). However, the Likert scale responses are translated into data points, which are then represented in an Apriori Table 1. Let a questionnaire be crafted and administered to a group of respondents, denoted as *R*, comprising a finite set of questions, *Q*, each with options scaled on a 5-point Likert scale (*V*). A relational data table, referred to as DB, is structured to accommodate responses from all respondents, which formed a dataset which is suitable for association rules mining within *Q*. Each entry in the relational data table DB is represented in Equation 1 as a tuple, denoted as $r_i \in R$, depicted as

$$r_i = \{(q_1, v_1), (q_2, v_2), \dots, (q_n, v_p)\}, \quad (1)$$

where r_i signifies the i^{th} respondent tuple. Here, $q_j \in Q$, denotes the set of questions within the questionnaire, and $v_k \in V$ represents the Likert point corresponding to each respondent for a given question. The tuple (q_j, v_k) signifies an atomic value within tuple r_i. For instance, Table 1 presents a hypothetical dataset extracted from a questionnaire comprising 7 questions and 8 respondents, each rated on a 5-point Likert scale $(S, A, N, \hat{A}, \hat{S})$.

Table 1: Sample Data

R.	q_1	q_2	q_3	q_4	q_5	q_6	q_7
r_1	N	Â	Â	N	A	Â	S
r_2	S	Â	S	Ŝ	N	Ŝ	S
r_3	A	N	S	A	Â	Â	A
r_4	Â	Ŝ	A	N	N	Ŝ	N
r_5	Â	N	N	A	Â	Â	A
r_6	A	S	Â	A	A	Â	S
r_7	N	A	Ŝ	N	Â	Â	S
r_8	N	A	Â	A	A	N	N

The keys for the Likert scale are defined in Table 2 [38], quantifying each respondent's response as a Likert value *V*, ranging from 1 to 5, alongside their corresponding keys.

Table 2: Likert Scale Keys

Likert Scale	Symbol	Value
Strongly Agree	S	5
Agree	A	4
Neutral	N	3
Disagree	Â	2
Strongly Disagree	Ŝ	1

3.1 PROBLEM DEFINITIONS

In order to effectively represent the 5-point Likert scale on the Apriori table, the following definitions were introduced:

Definition 1: Let $r_i \in R$ denote the set of all respondents, $q_j \in Q$ denote the set of all questions, and $v_k \in V$ denote the set of all Likert scale points corresponding to each question in Q. We define the information function in Equation 2 and 3 as follows;

$$\rho : R \times Q \to V, \tag{2}$$

such that

$$\rho(r_i, q_j) = v_k \tag{3}$$

where $i = 1, \ldots, m,\ j = 1, \ldots, n$ and $v_k = 1, \ldots, 5$.

Table 3: Likert Points Representation

R.	q_1	q_2	q_3	q_4	q_5	q_6	q_7
r_1	3	2	2	3	4	2	5
r_2	5	2	5	1	3	1	5
r_3	4	3	5	4	2	2	4
r_4	2	1	4	3	3	1	3
r_5	2	3	3	4	2	2	4
r_6	4	5	2	4	4	2	5
r_7	3	4	1	3	2	2	5
r_8	3	4	2	4	4	3	3

This function assigns an encoded Likert response (an integer between 1 and 5, inclusive) to a respondent for a given question. This is helpful in populating Table 3, where the sample data from Table 1 is translated into Likert key values based on Definition 1. These information functions are then depicted in Table 3. Additionally, Table 4 represents the Likert values selected by individual respondents for each question, providing a comprehensive view of the Likert scale responses.

Table 4: Data point on Apriori table

R	q_1	q_2	q_3	q_4	q_5	q_6	q_7
r_1	$\mu_{(1,3)}$	$\mu_{(2,2)}$	$\mu_{(3,2)}$	$\mu_{(4,3)}$	$\mu_{(5,4)}$	$\mu_{(6,2)}$	$\mu_{(7,5)}$
r_2	$\mu_{(1,5)}$	$\mu_{(2,2)}$	$\mu_{(3,5)}$	$\mu_{(4,1)}$	$\mu_{(5,3)}$	$\mu_{(6,1)}$	$\mu_{(7,5)}$
r_3	$\mu_{(1,4)}$	$\mu_{(2,3)}$	$\mu_{(3,5)}$	$\mu_{(4,4)}$	$\mu_{(5,2)}$	$\mu_{(6,2)}$	$\mu_{(7,4)}$
r_4	$\mu_{(1,2)}$	$\mu_{(2,1)}$	$\mu_{(3,4)}$	$\mu_{(4,3)}$	$\mu_{(5,3)}$	$\mu_{(6,1)}$	$\mu_{(7,3)}$
r_5	$\mu_{(1,2)}$	$\mu_{(2,3)}$	$\mu_{(3,3)}$	$\mu_{(4,4)}$	$\mu_{(5,2)}$	$\mu_{(6,2)}$	$\mu_{(7,4)}$
r_6	$\mu_{(1,4)}$	$\mu_{(2,5)}$	$\mu_{(3,2)}$	$\mu_{(4,4)}$	$\mu_{(5,4)}$	$\mu_{(6,2)}$	$\mu_{(7,5)}$
r_7	$\mu_{(1,3)}$	$\mu_{(2,4)}$	$\mu_{(3,1)}$	$\mu_{(4,3)}$	$\mu_{(5,2)}$	$\mu_{(6,2)}$	$\mu_{(7,5)}$
r_8	$\mu_{(1,3)}$	$\mu_{(2,4)}$	$\mu_{(3,2)}$	$\mu_{(4,4)}$	$\mu_{(5,4)}$	$\mu_{(6,3)}$	$\mu_{(7,3)}$

Definition 2: Let ∂_j represent the set of non-empty $\mu_{(q_j,v_k)}$ with Likert values $v_k \in V$ for question q_j. Formally in Equation 4 as,

$$\partial_j = \left\{ \mu_{(q_j,v_k)} \middle| v_k \in V, \mu_{(q_j,v_k)} \neq \varnothing \right\} \tag{4}$$

Note that there is total agreement among respondents if $|\partial_j| = 1$ and a total disagreement among respondents if $|\partial_j| = |V|$

Example 1:

$\partial_1 = \{\mu_{(1,2)}, \mu_{(1,3)}, \mu_{(1,4)}, \mu_{(1,5)}\}$.

$\partial_2 = \{\mu_{(2,2)}, \mu_{(2,3)}, \mu_{(2,1)}, \mu_{(2,4)}, \mu_{(2,5)}\}$.

$\partial_3 = \{\mu_{(3,1)}, \mu_{(3,2)}, \mu_{(3,3)}, \mu_{(3,4)}, \mu_{(3,5)}\}$.

$\partial_4 = \{\mu_{(4,1)}, \mu_{(4,3)}, \mu_{(4,4)}\}$.

$\partial_5 = \{\mu_{(5,2)}, \mu_{(5,3)}, \mu_{(5,4)}\}$.

This definition establishes the concept of ∂_j as the set of non-empty $\mu_{(q_j,v_k)}$ for each question q_j, illustrating the presence and diversity of Likert scale responses among respondents.

Definition 3: Let $\sigma = (q_j, v_k)$, where $q_j \in Q$ and $v_k \in V$. We define the set $(R)_\sigma \subseteq R$ that contains all respondents who responded with Likert value v_k for question q_j as shown in Equation 5 as;

$$(R)_\sigma = \{r_i | \rho(r_i, q_j) = v_k\} \tag{5}$$

Examples 2:

$(R)_{(1,3)} = \{r_1, r_7, r_8\}$,

The respondents r_1, r_7 and r_8 all responded NEUTRAL to question 1.

Likewise,

$(R)_{(4,4)} = \{r_3, r_5, r_6, r_8\}$

This shows that respondents r_3, r_5, r_6 and r_8 all responded AGREE to question 4.

Definition 4: A support count (S_c) is defined in Equation 6 as the number of respondents who responded with Likert value v_k for question q_j such that;

$$S_c = |(R)_\sigma| = \left|(R)_{(q_j,v_k)}\right| \tag{6}$$

Examples 3:

For instance, the support count S_c for $\sigma = (3,2)$ is expressed as follows;

$S_c = |(R)_{(3,2)}| = |\{r_1, r_6, r_8\}| = 3$

Similarly, when $\sigma = (7,5)$:

$S_c = |(R)_{(7,5)}| = |\{r_1, r_2, r_6, r_7\}| = 4$

Additionally, when σ consists of two or more itemsets, such as $((3,2),(7,5))$, then the support count is the number of respondents of the intersection of the itemsets:

$$S_c = \left|(R)_{(3,2)(7,5)}\right| = (r_1, r_6)_{((3,2),(7,5))} = \left|\{r_1, r_6\}\right| = 2$$

4. Experimental Method

This section describes the procedural steps taken in formulating the proposed algorithm, Extended Apriori Algorithm (Ext-AA). The framework involves a collection of questionnaires sourced from diverse respondents, which are stored subsequently in a central data warehouse as discussed in the previous section. The pruning phase of the Ext-AA framework employs a fundamental technique known as Common Question Attribute Pruning (CQAP), aimed at generating candidate item sets. Subsequently, the association rules segment utilizes newly defined support and confidence measures to derive robust and reasonable rules, thereby assessing the strength of each rule. The pseudo code in Algorithm 1,

Algorithm 1: Pseudocode for the Extended Apriori Algorithm (Ext-AA)

```
Input: Q ← Questions, V, R ← Respondent, min_sup, min_conf
Output: Array ← assoc_rules
QR ← user response
procedure APR()
procedure DPT_APR()
q ← {1, . . . , m}
v ← {1, 2, 3, 4, 5}
LIKERT_SCALE() ← v
for all i ∈ Q do
    for all j ∈ R do
        read ← LIKERT_SCALE(Q_i, R_j)
        dpt_Apr(i)(j) ← TRANSFORM(read)
    end for
end for
for all i ∈ dpt_Apr do
    for all j ∈ dpt_Apr(i) do
        // List each item in dpt_Apr without repetition
        L(q, v) ← LIST(dpt_Apr)
    end for
end for
k ← 1
C_k ← L(q, v)
L_k ← GETFREQ(C_k)
while L_k ≠ ∅ do
    C_{k+1} ← CQAP(L_k ⋈ L_k)
    k ← k + 1
end while
return ASSOC_RULES(L_k)
procedure GETFREQ(C_k)
for all c ∈ C_k do
    if S_c(c) < min_sup then
        C_k.remove(c)
    end if
end for
return C_k
```

outlines the Extended Apriori Algorithm (Ext-AA) for mining association rules from a 5-points Likert dataset. It starts by reading user responses and transforming them into a data structure *dpt_Apr*. Then, it iteratively generates candidate itemsets C_k and prunes them using the Common Question Attribute Pruning (CQAP) technique. Frequent itemsets satisfying the minimum support threshold are retained, and the process continues until all frequent item sets are generated. Finally, the Ext-AA returns the association rules derived from the frequent itemsets.

4.1 CQAP Technique

Common Question Attribute Pruning (**CQAP**) is a new approach designed specifically for Likert dataset for generating candidate itemsets on Apriori-like algorithm. Algorithm 2 outlines the Common Question Attribute Pruning (CQAP) approach, designed for Likert datasets, to generate candidate itemsets in an Apriori-like algorithm. It comprises three phases: support count, frequent item pruning, and joining phase [39]. The algorithm initializes data structures and calculates support counts for each item. It then iterates through the data to produce set of candidate items for the subsequent iteration. The joining phase prunes common question attributes and performs the joining operation on the frequent itemsets.

The resulting candidate sets are returned for further processing. At line 15 a procedure **getFreq** is called in order to generate the k+1 candidate sets for next iteration. The joining is called at line 16 which checks and prunes the common question attributes before performing the joining operations on L_k to itself. The joining procedure returns C_{k+1} at line 20 to line 28.

Algorithm 2: Pseudocode for the CQAP

```
Input: L_k, min sup
Output: C_{k+1} ← {}
1:  dpt_Apr ← load(L_k, min_sup)
2:  X(i, j) ← 0
3:  for all i ∈ dpt_Apr do
4:      for all j ∈ dpt_Apr(i) do
5:          C_i(X(i, j)) ← datapoint_Apr(i)(j)
6:      end for
7:  end for
8:  S_c ← {}
9:  for all x ∈ C_i do
10:     S_c(x) ← COUNT_OCCURENCE(C_i(x), C_i)
11: end for
12: k ← 1
13: for all i in range(k, dpt_Apr.length) do
14:     for all j in range(k, dpt_Apr.length) do
15:         L_{k+1} ← GETFREQ(C_{k+1})
16:         C_{k+i} ← JOIN(L_k)
17:     end for
18:     k ← k + 1
19: end for
20: procedure JOIN(L_k)
21:     for all Q(u, v) ∈ L_k do
22:         for all W(i, j) ∈ L_k do
23:             if Q(u, v) ∧ W(i, j) joinable, if and only if u ≠ i then
24:                 C_{k+i} ←    {Q( u, v) ⋈ W (i, j)} :(u, i) ← 1, 2, . . . ,
    dptApr.columns
25:                 C_{k+1} ← append(C_{k+i}) : (j, v) ← 1, 2, . . . , dpt_Apr.length
26:             else
27:                 pass
28:             end if
29:         end for
30:     end forreturn C_{k+1}
31: end procedure
```

4.2 Illustrative steps for the proposed Ext-AA

Step 1: *Generating the Candidate itemset C_1 and Support*.
To generate the candidate itemset C_1 and the support count (S_c) of each data point $\left(\mu_{(q_i,v_k)}\right)$, scan through Table 4 using Definition 4. This involves determining all the possible data points σ_i with their corresponding (S_c). Each σ_i is called the candidate set C_1 and is represented in Table 5.

The process can be outlined as follows:

1. **Scan through Table 4**: Iterate through each respondent's responses to identify all unique question-Likert value pairs (q_i, v_k).
2. **Apply Definition 4**: For each identified pair (q_i, v_k), count the number of respondents who selected this particular Likert value for the given question. This count is the support count S_c.
3. **Generate Candidate Set C_1**: Compile all unique (q_i, v_k), pairs into a list, where each pair is

associated with its support count S_c. This list forms the candidate set C_1.

4. **Populate Table 5**: Record each unique (q_i, v_k), pair along with its support count S_c in Table 5.

Example 4:
Suppose Table 4 contains responses for 7 questions from 8 respondents. By applying Definition 4, you might find pairs such as:

- $(q_1, 3)$ with a support count of 3
- $(q_2, 3)$ with a support count of 2
- $(q_3, 2)$ with a support count of 3

These pairs will be listed in Table 5 along with their respective support counts, forming the candidate set C_1, as shown in Algorithm 3.

```
Algorithm 3: Generate C1 and Support Count
Input: Response Table (Table 4)
Output: Candidate Set C1 with Support Counts (Table 5)
Initialize an empty dictionary C1
for all respondent r in Table 4 do
    for all question q_i do
        Let v_k be the Likert value chosen by respondent r for question q_i
        if (q_i, v_k) not in C1 then
            Add (q_i, v_k) to C1 with count 1
        else
            Increment the count of (q_i, v_k) in C1 by 1
        end if
    end for
end for
Output C1 as Table 5
```

The support count of each item in C_1 from Algorithm 3 are presented in Table 5. By following these steps, the candidate set C_1 and their corresponding support counts S_c are generated and recorded, providing the foundation for further analysis and pruning in the subsequent steps.

Table 5: C_1 Candidate

C_1	(S_c)	C_1	(S_c)
$\mu_{(1,2)}$	2	$\mu_{(3,5)}$	2
$\mu_{(1,3)}$	3	$\mu_{(4,1)}$	1
$\mu_{(1,4)}$	2	$\mu_{(4,3)}$	3
$\mu_{(1,5)}$	1	$\mu_{(4,4)}$	4
$\mu_{(2,1)}$	1	$\mu_{(5,2)}$	3
$\mu_{(2,2)}$	2	$\mu_{(5,3)}$	2
$\mu_{(2,3)}$	2	$\mu_{(5,4)}$	3
$\mu_{(2,4)}$	2	$\mu_{(6,1)}$	2
$\mu_{(2,5)}$	1	$\mu_{(6,2)}$	5
$\mu_{(3,1)}$	1	$\mu_{(6,3)}$	1
$\mu_{(3,2)}$	3	$\mu_{(7,3)}$	2
$\mu_{(3,3)}$	1	$\mu_{(7,4)}$	2
$\mu_{(3,4)}$	1	$\mu_{(7,5)}$	4

Step 2: Generating L_1 by from the Candidate Set C_1
To generate L_1, the candidate set C_1 is prune with respect to the threshold value. The support count S_c of each data point σ_i in C_1 is compared against the specified minimum support. Let's set the minimum support $\mathrm{Sup}_{\min}$ to 2. This implies that the support threshold is calculated as follows:

$$Support\ Threshold = \frac{\mathrm{Sup}_{min}}{Total\ Respondents} \times 100$$

$$= \frac{2}{8} * 100 = 25\%$$

Pruning Process:

1. **Compare Support Counts:** For each candidate C_1, check if the support count $S_c(\sigma_i)$ is greater than or equal $\mathrm{Sup}_{\min}$.
2. **Select Frequent Items:** Candidates with $S_c(\sigma_i) \geq 2$ are selected. The candidates form the frequent itemset L_1.
3. **Populate Table 6:** Record the pruned candidates and their support counts in Table 6 which represents L_1.

Table 6: L_1 itemset

L_1	S_c	L_1	S_c
$\mu_{(1,2)}$	2	$\mu_{(4,4)}$	4
$\mu_{(1,3)}$	3	$\mu_{(5,2)}$	3
$\mu_{(1,4)}$	2	$\mu_{(5,3)}$	2
$\mu_{(2,2)}$	2	$\mu_{(5,4)}$	3
$\mu_{(2,3)}$	2	$\mu_{(6,1)}$	2
$\mu_{(2,4)}$	2	$\mu_{(6,2)}$	5
$\mu_{(3,2)}$	3	$\mu_{(7,3)}$	2
$\mu_{(3,5)}$	2	$\mu_{(7,4)}$	2
$\mu_{(4,3)}$	3	$\mu_{(7,5)}$	4

Definition 5: Let $X \subseteq Q$, and $\Im$ be a *joining function* defined on a set of candidates μ_i such that $q_j, q_t \in X$, where $j < t$, $X \geq Sup_{\min}$ and for all $v_p, v_u \in V$, then $(\mu_i)_\Im = \mu_i \bowtie \mu_i$, if and only if for all $q_j, q_t \in Q$, as shown in Equation 7 and 8.

$$\left\{\mu_{(q_j,v_k)}, \mu_{(q_t,v_p)}\right\} = \begin{cases} join, j \neq t \\ prune, j = t \end{cases} \tag{7}$$

$$(\mu_i)_\Im = \left\{\mu_{(q_j,v_p)}, \mu_{(q_t,v_u)}\right\} \text{ is joinable.} \tag{8}$$

Where q_j, q_t denote *uncommon question attribute* of $\mu_{i(k)}$ and $\mu_{i_{(k+1)}}$ respectively.

Step 3: Generation of C_2 Candidate Set
The C_2 candidate set is generated by joining L_1 candidate set to itself (i.e., $C_2 = \mu_i \bowtie \mu_i$). This joining process involves two major cases:

Case 1: No Common Question Attribute joining: Based on definition 5, some candidates in L_1 will not be joined if they consist of *common question attribute.* For instance, if one respondent selects $N = 3$ for q_1 and another respondent selects $A = 4$ for same q_1, then,

$\mu_{i_{(k)}}: \{q_1, 3\}$ *and* $\mu_{(k+1)}: \{q_1, 4\}$

For the same question q_1, having two different Likert point $v = 3$ and $v = 4$ in the same column, are *impermissible* to join. Therefore, by Definition 5, some of the L_1 candidates set (consisting of the same question on the same column) are pruned before generating the C_2 candidate set in the joining process. This approach helps in saving time by avoiding the generation of irrelevant itemsets that may not be frequent in subsequent steps.

Case 2: Apriori Monotonic Principle: The second case relies on the Apriori monotonic principle, which states that if a subset is infrequent, then its superset is also infrequent. Based on this principle, all infrequent data points in the Apriori Table 4 are pruned. The table size is reduced and updated, as shown in Table 7. These cases were introduced to reduce the size of the database and to generate a smaller number of candidate sets (relevant candidate sets) whose S_c value is at least the given $Sup_{min}(2)$. Since no Common Question Attribute of two or more *data points* are allowed to be joined together based on definition 5, Table 8 shows C_2 candidate set which involves only a set of *uncommon question attribute* and *frequent* itemset. In Table 7, the following are some of the *datapoints* $(L_1 \bowtie L_1)$ that qualified to generate C_2 ; $\{\mu_{(1,2)}, \mu_{(2,2)}\}, \{\mu_{(1,2)}, \mu_{(2,3)}\}, \{\mu_{(1,2)}, \mu_{(2,4)}\}, \{\mu_{(1,2)}, \mu_{(3,2)}\}$ $\{\mu_{(1,2)}, \mu_{(3,5)}\}, \{\mu_{(1,2)}, \mu_{(4,3)}\}, \{\mu_{(1,2)}, \mu_{(4,4)}\}, \{\mu_{(1,2)}, \mu_{(5,2)}\}$ $\{\mu_{(1,2)}, \mu_{(5,3)}\}, \{\mu_{(1,2)}, \mu_{(5,4)}\}$.

However, the following data points $\mu_{(2,2)}, \mu_{(2,4)}, \mu_{(3,2)}, \mu_{(3,5)}$ *and* $\mu_{(5,4)}$ cannot be joined with $\mu_{(1,2)}$ because they do not exist in rows r_4 and r_5 which contain $\mu_{(1,2)}$ in the updated Table 7.

Table 7: Updated Apriori Table for C_2 Candidate set

R	q_1	q_2	q_3	q_4	q_5	q_6	q_7
r_1	$\mu_{(1,3)}$	$\mu_{(2,2)}$	$\mu_{(3,2)}$	$\mu_{(4,3)}$	$\mu_{(5,4)}$	$\mu_{(6,2)}$	$\mu_{(7,5)}$
r_2	–	$\mu_{(2,2)}$	$\mu_{(3,5)}$	–	$\mu_{(5,3)}$	$\mu_{(6,1)}$	$\mu_{(7,5)}$
r_3	$\mu_{(1,4)}$	$\mu_{(2,3)}$	$\mu_{(3,5)}$	$\mu_{(4,4)}$	$\mu_{(5,2)}$	$\mu_{(6,2)}$	$\mu_{(7,4)}$
r_4	$\mu_{(1,2)}$	–	–	$\mu_{(4,3)}$	$\mu_{(5,3)}$	$\mu_{(6,1)}$	$\mu_{(7,3)}$
r_5	$\mu_{(1,2)}$	$\mu_{(2,3)}$	–	$\mu_{(4,4)}$	$\mu_{(5,2)}$	$\mu_{(6,2)}$	$\mu_{(7,4)}$
r_6	$\mu_{(1,4)}$	–	$\mu_{(3,2)}$	$\mu_{(4,4)}$	$\mu_{(5,4)}$	$\mu_{(6,2)}$	$\mu_{(7,5)}$
r_7	$\mu_{(1,3)}$	$\mu_{(2,4)}$	–	$\mu_{(4,3)}$	$\mu_{(5,2)}$	$\mu_{(6,2)}$	$\mu_{(7,5)}$
r_8	$\mu_{(1,3)}$	$\mu_{(2,4)}$	$\mu_{(3,2)}$	$\mu_{(4,4)}$	$\mu_{(5,4)}$	–	$\mu_{(7,3)}$

Furthermore, based on these cases stated above, C_2 $(L_1 \bowtie L_1)$ candidates set were generated, as shown in Table 8.

Table 8: C_2 -Candidate set

C_2-candidate	S_c	C_2-candidate	S_c	C_2-candidate	S_c
$\mu_{(1,2)}, \mu_{(2,3)}$	1	$\mu_{(1,4)}, \mu_{(3,2)}$	1	$\mu_{(2,3)}, \mu_{(7,4)}$	2
$\mu_{(1,2)}, \mu_{(4,3)}$	1	$\mu_{(1,4)}, \mu_{(3,5)}$	1	$\mu_{(2,4)}, \mu_{(3,2)}$	1
$\mu_{(1,2)}, \mu_{(4,4)}$	1	$\mu_{(1,4)}, \mu_{(4,4)}$	2	$\mu_{(2,4)}, \mu_{(4,3)}$	1
$\mu_{(1,2)}, \mu_{(5,2)}$	1	$\mu_{(1,4)}, \mu_{(5,2)}$	1	$\mu_{(2,4)}, \mu_{(4,4)}$	1
$\mu_{(1,2)}, \mu_{(5,3)}$	1	$\mu_{(1,4)}, \mu_{(5,4)}$	1	$\mu_{(2,4)}, \mu_{(5,2)}$	1
$\mu_{(1,2)}, \mu_{(6,1)}$	1	$\mu_{(1,4)}, \mu_{(6,2)}$	2	$\mu_{(2,4)}, \mu_{(5,4)}$	1
$\mu_{(1,2)}, \mu_{(6,2)}$	1	$\mu_{(1,4)}, \mu_{(7,4)}$	1	$\mu_{(2,4)}, \mu_{(6,2)}$	1
$\mu_{(1,2)}, \mu_{(7,3)}$	1	$\mu_{(1,4)}, \mu_{(7,5)}$	1	$\mu_{(2,4)}, \mu_{(7,3)}$	1
$\mu_{(1,2)}, \mu_{(7,4)}$	1	$\mu_{(2,2)}, \mu_{(3,2)}$	1	$\mu_{(2,4)}, \mu_{(7,5)}$	1
$\mu_{(1,3)}, \mu_{(2,2)}$	1	$\mu_{(2,2)}, \mu_{(3,5)}$	1	$\mu_{(3,2)}, \mu_{(4,3)}$	1
$\mu_{(1,3)}, \mu_{(2,4)}$	2	$\mu_{(2,2)}, \mu_{(4,3)}$	1	$\mu_{(3,2)}, \mu_{(4,4)}$	2
$\mu_{(1,3)}, \mu_{(3,2)}$	2	$\mu_{(2,2)}, \mu_{(5,3)}$	1	$\mu_{(3,2)}, \mu_{(5,4)}$	3
$\mu_{(1,3)}, \mu_{(4,3)}$	2	$\mu_{(2,2)}, \mu_{(5,4)}$	1	$\mu_{(3,2)}, \mu_{(6,2)}$	2
$\mu_{(1,3)}, \mu_{(4,4)}$	1	$\mu_{(2,2)}, \mu_{(6,1)}$	1	$\mu_{(3,2)}, \mu_{(7,3)}$	1
$\mu_{(1,3)}, \mu_{(5,2)}$	1	$\mu_{(2,2)}, \mu_{(6,2)}$	1	$\mu_{(3,2)}, \mu_{(7,5)}$	2
$\mu_{(1,3)}, \mu_{(5,4)}$	2	$\mu_{(2,2)}, \mu_{(7,5)}$	2	$\mu_{(4,3)}, \mu_{(5,2)}$	1
$\mu_{(1,3)}, \mu_{(6,2)}$	2	$\mu_{(2,3)}, \mu_{(3,5)}$	1	$\mu_{(4,3)}, \mu_{(5,3)}$	1
$\mu_{(1,3)}, \mu_{(7,3)}$	1	$\mu_{(2,3)}, \mu_{(4,4)}$	2	$\mu_{(4,3)}, \mu_{(6,1)}$	1
$\mu_{(1,3)}, \mu_{(7,5)}$	2	$\mu_{(2,3)}, \mu_{(5,2)}$	2	$\mu_{(4,3)}, \mu_{(6,2)}$	2
$\mu_{(1,4)}, \mu_{(2,3)}$	1	$\mu_{(2,3)}, \mu_{(6,2)}$	2	$\mu_{(4,3)}, \mu_{(7,3)}$	1

Step 4: Generate L_2 itemset from Table 8 by pruning the C_2 -candidate set whose support count (S_c < 2) is less than the minimum support value. Therefore, L_2 itemset are the list of frequent itemset in C_2 -candidate set as shown in Table 9.

Step 5: However, based on cases stated in step 3, the L_2 itemset in Table 9 was joined $(L_2 \bowtie L_2)$ to itself and 25 C_3-candidates set were generated of which only 10 C_3-candidates were frequent to make L_3 itemset as shown in Table 10.

Step 6: To generate C_4 candidate set from Table 10, repeat step 5 (i.e., $(L_3 \bowtie L_3)$) on L_3 itemset. The Table 11 depicts 8 C_4-candidates generated after the joining process.

Step 7: Repeat step 4 to determine L_4 itemset from Table 11. Out of 8 C_4-candidates generated, only 1of its candidate is frequent to generate L_4 itemset as shown in Table 12. Therefore, $\{\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}\}$ itemset is the only frequent item in L_4 required to generate association rules for the sample data.

Table 9: L_2 Itemset

L_2 candidate	S_c	L_2 candidate	S_c
$\mu_{(1,3)}, \mu_{(2,4)}$	2	$\mu_{(3,2)}, \mu_{(6,2)}$	2
$\mu_{(1,3)}, \mu_{(3,2)}$	2	$\mu_{(3,2)}, \mu_{(7,5)}$	2
$\mu_{(1,3)}, \mu_{(4,3)}$	2	$\mu_{(4,3)}, \mu_{(6,2)}$	2
$\mu_{(1,3)}, \mu_{(5,4)}$	2	$\mu_{(4,3)}, \mu_{(7,5)}$	2
$\mu_{(1,3)}, \mu_{(6,2)}$	2	$\mu_{(4,4)}, \mu_{(5,2)}$	2
$\mu_{(1,3)}, \mu_{(7,5)}$	2	$\mu_{(4,4)}, \mu_{(5,4)}$	2
$\mu_{(1,4)}, \mu_{(4,4)}$	2	$\mu_{(4,4)}, \mu_{(6,2)}$	3
$\mu_{(1,4)}, \mu_{(6,2)}$	2	$\mu_{(4,4)}, \mu_{(7,4)}$	2
$\mu_{(2,2)}, \mu_{(7,5)}$	2	$\mu_{(5,2)}, \mu_{(6,2)}$	3
$\mu_{(2,3)}, \mu_{(4,4)}$	2	$\mu_{(5,2)}, \mu_{(7,4)}$	2
$\mu_{(2,3)}, \mu_{(5,2)}$	2	$\mu_{(5,3)}, \mu_{(6,1)}$	2
$\mu_{(2,3)}, \mu_{(6,2)}$	2	$\mu_{(5,4)}, \mu_{(6,2)}$	2
$\mu_{(2,3)}, \mu_{(7,4)}$	2	$\mu_{(5,4)}, \mu_{(7,5)}$	2
$\mu_{(3,2)}, \mu_{(4,4)}$	2	$\mu_{(6,2)}, \mu_{(7,4)}$	2
$\mu_{(3,2)}, \mu_{(5,4)}$	3	$\mu_{(6,2)}, \mu_{(7,5)}$	3

Table 10: L_3 Candidate Set

L_3	S_c
$\mu_{(3,2)}, \mu_{(4,4)}, \mu_{(5,4)}$	2
$\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(6,2)}$	2
$\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(7,5)}$	2

$\mu_{(3,2)}, \mu_{(6,2)}, \mu_{(7,5)}$	2
$\mu_{(4,3)}, \mu_{(6,2)}, \mu_{(7,5)}$	2
$\mu_{(4,4)}, \mu_{(5,2)}, \mu_{(6,2)}$	2
$\mu_{(4,4)}, \mu_{(5,2)}, \mu_{(7,4)}$	2
$\mu_{(4,4)}, \mu_{(6,2)}, \mu_{(7,4)}$	2
$\mu_{(5,2)}, \mu_{(6,2)}, \mu_{(7,4)}$	2
$\mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}$	2

Table 11: C_4 Candidate set

C_4	S_c
$\mu_{(3,2)}, \mu_{(4,3)}, \mu_{(5,4)}, \mu_{(6,2)}$	1
$\mu_{(3,2)}, \mu_{(4,3)}, \mu_{(5,4)}, \mu_{(7,5)}$	1
$\mu_{(3,2)}, \mu_{(4,4)}, \mu_{(5,4)}, \mu_{(6,2)}$	1
$\mu_{(3,2)}, \mu_{(4,4)}, \mu_{(5,4)}, \mu_{(7,5)}$	1
$\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}$	2
$\mu_{(4,3)}, \mu_{(5,2)}, \mu_{(6,2)}, \mu_{(7,5)}$	1
$\mu_{(4,3)}, \mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}$	1
$\mu_{(4,4)}, \mu_{(5,2)}, \mu_{(6,2)}, \mu_{(7,4)}$	1

Table 12: L_4 itemset

L_4 itemset	**S_c**
$\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}$	2

4.2 Association Rule Generation

In data mining, Association rules generation is the final step which reveals the relationships between the data attributes in a particular research domain. Likert scale dataset have been successfully mined on Ext-AA thereby extending the capacity of the Apriori from binary mining process to 5-point scale value mining using CQAP technique. However, from Table 12, only one L_4 -itemset is generated for the required association rules. The $\{\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}\}$ -item comprises of 14 proper subsets which represent the association rules generated. The followings are the proper subset of

$\{\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}\}$ are $\{(\mu_{(3,2)}), (\mu_{(5,4)}), (\mu_{(6,2)}), (\mu_{(7,5)})\}, \{(\mu_{(3,2)}, \mu_{(5,4)}), (\mu_{(3,2)}, \mu_{(6,2)}), (\mu_{(3,2)}, \mu_{(7,5)}), (\mu_{(5,4)}, \mu_{(6,2)}), (\mu_{(5,4)}, \mu_{(7,5)}), (\mu_{(6,2)}, \mu_{(7,5)})\}, \{(\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(6,2)}), (\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(7,5)}), (\mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}), (\mu_{(3,2)}, \mu_{(6,2)}, \mu_{(7,5)})\}$.

Definition 6: Let $\boldsymbol{R}$ be a set of respondents in the database, $\boldsymbol{X} \subseteq \boldsymbol{Q}$, be a set of questions, with its corresponding Likert value $\boldsymbol{v}$ of 5-points scale such that $\boldsymbol{v} \in \boldsymbol{V}$, We define support (Sup) in Equation 9 and 10 as follows;

$$Sup\{\mu_{(q_1,v)}\} = \frac{|\mu_{(q_1,v)}|}{R} \quad (9)$$

$$Sup(\mu_{(q_1,v)} \cup \mu_{(q_2,v)} \cup, ..., \mu_{(q_n,v)}) = \frac{|\mu_{(q_1,v)} \cup, ..., \cup \mu_{(q_n,v)}|}{R} \quad (10)$$

where $q_1, q_2, \ldots, q_n \in X$ and $v = [1,5]$, is the Likert' response value to a question q_i for each pair $\mu_{(q_i,v)}$.

Therefore,

$$Sup(\mu_{(3,2)} \cup \mu_{(4,4)} \cup, \mu_{(5,4)}) = \frac{|\mu_{(3,2)} \cup \mu_{(4,4)} \cup \mu_{(5,4)}|}{8} = \frac{2}{8}$$

$= 25\%$

Definition 7: Let $\boldsymbol{A} \subset \boldsymbol{Q}$, be an antecedent question-value pair such that, $\boldsymbol{A} = \{\mu_{(q_a,v)}\}$, and $\boldsymbol{C} \subset \boldsymbol{Q}$, be a consequent question-value pair such that, $\boldsymbol{C} = \{\mu_{(q_c,v)}\}$, with their corresponding Likert value $(\boldsymbol{v})$ of 5-points such that $\boldsymbol{v} \in \boldsymbol{V}$, We define confidence, (*Conf.*) in Equation 11 and 12 as;

$$Conf\,(A \rightarrow C) = \frac{Sup(A \cup C)}{Sup(A)} = \frac{|A \cup C|}{|A|}, \quad (11)$$

$$Conf\,(\mu_{(q_a,v)} \rightarrow \mu_{(q_c,v)}) = \frac{|\mu_{(q_a,v)} \cup \mu_{(q_c,v)}|}{|\mu_{(q_a,v)}|} \quad (12)$$

Where $a \neq c$ and for all $q_a, q_c \in Q$.

Therefore, from definition 6 and definition 7, the L_4 itemset generates 14 distinctive rules in which the strength of individual rule generated was determined by the specified minimum confidence value (70%). The following rules are some of the 14 rules computed;

Rule a: $(\mu_{(3,2)}) \rightarrow (\mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)})$:

$((\mu_{(3,2)}\mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}) - (\mu_{(3,2)}))$

$$Sup(\mu_{(3,2)}) = \frac{|\mu_{(3,2)}|}{R} = \frac{3}{8} = 0.375 = 37.5\%$$

$$Conf\left((\mu_{(3,2)} \rightarrow (\mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)})\right)$$

$$= \frac{Sup(\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)})}{Sup(\mu_{(3,2)})} = \frac{2}{3} \approx 0.667 = 66.7\%$$

Rule b: $(\mu_{(5,4)}, \mu_{(6,2)}) \rightarrow (\mu_{(3,2)}, \mu_{(7,5)})$:

$((\mu_{(3,2)}\mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}) - (\mu_{(5,4)}, \mu_{(6,2)}))$

$$Sup(\mu_{(5,4)}, \mu_{(6,2)}) = \frac{|\mu_{(5,4)}, \mu_{(6,2)}|}{R} = \frac{2}{8} = 0.250 = 25.0\%$$

$$Conf\left((\mu_{(5,4)}, \mu_{(6,2)}) \rightarrow (\mu_{(3,2)}, \mu_{(7,5)})\right)$$

$$= \frac{Sup(\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)})}{Sup(\mu_{(5,4)}, \mu_{(6,2)})} = \frac{2}{2} = 1.00 = 100\%$$

Rule c: $(\mu_{(3,2)}, \mu_{(6,2)}, \mu_{(7,5)}) \rightarrow (\mu_{(5,4)})$:

$((\mu_{(3,2)}\mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}) - (\mu_{(3,2)}, \mu_{(6,2)}, \mu_{(7,5)}))$

$$Sup(\mu_{(3,2)}, \mu_{(6,2)}, \mu_{(7,5)}) = \frac{|Sup(\mu_{(3,2)}, \mu_{(6,2)}, \mu_{(7,5)})|}{R} = \frac{2}{8}$$

$= 0.250 = 25.0\%$

$$Conf\left((\mu_{(3,2)}, \mu_{(6,2)}, \mu_{(7,5)}) \rightarrow (\mu_{(5,4)})\right) =$$

$$\frac{Sup(\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)})}{Sup(\mu_{(3,2)}, \mu_{(6,2)}, \mu_{(7,5)})} = \frac{2}{2} = 1.00 = 100.0\%$$

5. Results and Discussion

In association rule mining, the generation of numerous rules is a common outcome; however, not all of these rules are relevant or useful for decision-making [45]. To address this issue, a minimum confidence ($Conf_{min}$) was employed to identify and retain only the strongest rules. This threshold serves as a pruning mechanism, filtering out rules that do not meet the user-specified confidence level. For the purpose of this study, the $Conf_{min}$ was set to 70%, meaning that any rule with a confidence lower than this threshold was discarded, while those with a confidence of 70% or higher were considered strong and retained for further analysis.

Table 13 illustrates the set of rules generated by the L3 itemset, where each rule's support count meets or exceeds the specified minimum support (Sup_{min}) of 2. The confidence levels of these rules were compared against the $Conf_{min}$ threshold, and only the rules that met or exceeded 70% confidence were considered significant.

Due to the binary nature of the standard Apriori algorithm, it was necessary to preprocess the sample data by discretizing and converting it into a transaction dataset before applying the algorithm in the existing research in [44]. The standard Apriori algorithm operates using binary values: 0 indicates an item was not purchased, while 1 indicates that it was purchased. To align the 5-point Likert scale data with this binary framework, a min-max normalization technique was employed. This technique scaled the dataset so that all values fell within the range [0, 1], where 0 represents the average mean and 1 represents the standard deviation. Following this normalization, data points with values below 0.5 were coded as 0, and those with values of 0.5 or higher were coded as 1. This conversion effectively transformed the Likert scale data into a transactional format suitable for analysis by the Apriori algorithm. A coded value of 0 indicated a negative response from the respondent, while a coded value of 1 indicated a positive response.

The application of the standard Apriori algorithm to this binary dataset resulted in the generation of 135 rules, as shown in Table 13. In contrast, the proposed Extended Apriori Algorithm (Ext-AA) generated only 14 rules, as shown in Table 14. Remarkably, 8 of these rules demonstrated a confidence level of 1.00, indicating their robustness and reliability. The significant reduction in the number of rules generated by Ext-AA compared to the standard Apriori highlights its efficiency in identifying only the most relevant and strong rules, thus reducing the complexity and enhancing the interpretability of the results.

This comparison highlights the effectiveness of the Ext-AA in focusing on the most meaningful patterns within the data, making it a more powerful tool for association rule mining, particularly in cases where the dataset involves ordinal data like the Likert scale. The application of the Ext-AA not only streamlined the rule generation process but also ensured that the strongest and most reliable rules were identified, facilitating more accurate and actionable insights.

Rule 4: $Conf(2 \rightarrow 6) = 83.3\%$ with $Sup = 62.5\%$. This rule means that 62.5% of Respondents who respond POSITIVE to Q_2 have 83.3% confidence to respond POSITIVE to Q_6.

Rule 133: $Conf(3,6 \rightarrow 2,4,7) = 100.0\%$ with $Sup = 25.0\%$. This rule means that 25% of Respondents who respond POSITIVE to Q_3 and Q_6 have 100% confidence to respond POSITIVE to Q_2, Q_4, Q_7. However, based on the specified $\boldsymbol{Conf_{min}}$ value which is 70%, only 8 rules out of 14 rules generated met the minimum confidence value and were considered as the strongest rules
To interpret the rules generated by the proposed Ex-AA as shown in Table 14, some of the strongest rules are represented as follow;

Rule 7: $Conf(\mu_{(3,2)}, \mu_{(7,5)} \rightarrow \mu_{(5,4)}, \mu_{(6,2)}) = 100\%$ with $Sup = 25\%$. This rule means that 25% of Respondents who DISAGREE to Q_3 and STRONGLY AGREE to Q_7 will have 100% confidence to Agree to Q_5 and DISAGREE to Q_6.

Rule 13: $Conf(\mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)} \rightarrow \mu_{(3,2)}) = 100\%$ with $Sup = 25\%$. This rule means that 25% of Respondents who AGREE to Q_5, DISAGREE to Q_6 and STRONGLY AGREE to Q_7 will have 100% confidence to DISAGREE to Q_3.

Table 13: Association rules measurement of the standard Apriori [44]

S/No.	antecedent	consequent	support	confidence
0	(1)	(7)	0.375	1.0000
1	(4)	(2)	0.625	1.0000
2	(2)	(4)	0.625	0.7143
3	(6)	(2)	0.625	1.0000
4	(2)	(6)	0.625	0.8333
...	...	...	...	...
133	(6, 3)	(4, 2, 7)	0.250	1.0000
134	(3, 2)	(6, 4, 7)	0.250	1.0000

Finally, the performance of the proposed Ext-AA was compared against the enhanced Apriori in [44], the Table 15 depicts the comparative analysis of the performance of the standard Apriori algorithm and the proposed Extended Apriori Algorithm (Ext-AA) across several key metrics. The Apriori algorithm generates 134 rules, while the Ext-AA produces only 14 which is 89.63% reduction, this demonstrates a significant reduction in the number of rules generated which enhance interpretability, and also highlights the efficiency of Ext-AA in focusing on the most relevant patterns. The Apriori algorithm operates on a binary scale (Yes = 1, No = 0), whereas Ext-AA retains the ordinal nature of the data, handling a 5-point Likert scale ranging from Strongly Agree (5) to Strongly Disagree (1). richness of the original data and generating more meaningful insights. Both algorithms used the same minimum support (Sup_{min} = 2) and minimum confidence ($Conf_{min}$ = 70%), but Ext-AA's ability to process ordinal data makes it more effective in preserving the richness of the original data and generating more meaningful insights.

Table 14: Association rules measurement of the proposed Ext-AA

S/N	antecedents	consequents	confidence
1.	$\mu_{(3,2)}$	$\mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}$	0.667
2.	$\mu_{(5,4)}$	$\mu_{(3,2)}, \mu_{(6,2)}, \mu_{(7,5)}$	0.667
3.	$\mu_{(6,2)}$	$\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(7,5)}$	0.400
4.	$\mu_{(7,5)}$	$\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(6,2)}$	0.500
5.	$\mu_{(3,2)}, \mu_{(5,4)}$	$\mu_{(6,2)}, \mu_{(7,5)}$	0.667
6.	$\mu_{(3,2)}, \mu_{(6,2)}$	$\mu_{(5,4)}, \mu_{(7,5)}$	1.000
7.	$\mu_{(3,2)}, \mu_{(7,5)}$	$\mu_{(5,4)}, \mu_{(6,2)}$	1.000
8.	$\mu_{(5,4)}, \mu_{(6,2)}$	$\mu_{(3,2)}, \mu_{(7,5)}$	1.000
9.	$\mu_{(5,4)}, \mu_{(7,5)}$	$\mu_{(3,2)}, \mu_{(6,2)}$	1.000
10.	$\mu_{(6,2)}, \mu_{(7,5)}$	$\mu_{(3,2)}, \mu_{(5,4)}$	0.667
11.	$\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(6,2)}$	$\mu_{(7,5)}$	1.000
12.	$\mu_{(3,2)}, \mu_{(5,4)}, \mu_{(7,5)}$	$\mu_{(6,2)}$	1.000
13.	$\mu_{(5,4)}, \mu_{(6,2)}, \mu_{(7,5)}$	$\mu_{(3,2)}$	1.000
14.	$\mu_{(3,2)}, \mu_{(6,2)}, \mu_{(7,5)}$	$\mu_{(5,4)}$	1.000

6. Conclusion and Future Scope

In this study, we have introduced the Common Question Attribute Pruning (CQAP) technique within the Apriori algorithm framework for mining 5-point Likert scale data. This approach aims to generate a smaller and more meaningful set of candidate itemsets with the corresponding association rules. The proposed CQAP technique extends the standard Apriori algorithm by directly accommodating 5-point Likert scale data, thus avoiding the need for discretization into traditional Boolean data (0s and 1s). We have redefined the support and confidence measures to better determine the strongest association rules for Likert data.

Our demonstration on a sample dataset shows the effectiveness of the CQAP technique in Ext-AA, which maintains the generality of respondents' perceptions without the need for discretization. The results indicate that the proposed Ext-AA algorithm reduced rules generated by 89.63%, this highlight how easy the rules can be interpreted compared to the standard Apriori algorithm, which primarily focus on only positive responses. This substantial reduction in the number of rules highlights the efficiency of the Ext-AA algorithm in accurately and effectively representing respondents' opinions or perceptions within a research domain. In future work, the joining phase will be optimized by using column-wise itemset merging technique which was presented in [22] and adaptive minimum support threshold technique will be adopted as presented in [43] to determine the optimal minimum support threshold value for the mining procedure in order to strike the balance between rule's specificity and generality in a research domain.

Data Availability

Data sharing is not applicable

Conflict of Interest

There is no conflict of interest in this study.

Funding Source

There is no funding source exists for this study.

Authors' Contributions

The Author-1 drafted the research work, while Author-2 and Author-3 reviewed, and approved the manuscript for publication.

Acknowledgements

Our unreserve gratitude goes to the staff of the department of computer science and computer engineering Ahmadu Bello University, Zaria, for their contributions and supports towards the success of this research.

References

[1] K. Baffour, C. Osei-Bonsu, A. Adekoya, "A Modified Apriori Algorithm for Fast and Accurate Generation of Frequent Item Sets". International Journal of Scientific & Technology Research (IJSTR), Vol.**6**, pp.**169-173**, **2017**.

[2] R. Z. Niu, G. Fang, D. Zeng, H. Yuan, "An improved Apriori algorithm based on influence weight". *In Third International Conference on Computer Vision and Pattern Analysis (ICCPA 2023)*, Vol.**12754**, pp.**673 679, 2023**.SPIE.

[3] X. Wu, Y. Xiao, A. Liu, "Application of improved Apriori Algorithm in Innovation and Entrepreneurship Engineering Education Platform". *Scalable Computing: Practice and Experience*, Vol.**24**, No.**3**, pp.**609-620, 2023**.

[4] E.S. Harpe, D. Pharm, "How to analyze Likert and other rating scale data". *Curr. Pharm. Teach. Learn*. Vol.**7**, pp.**836-850**, **2015**.

[5] U. Sivarajah, M. Kamal, Z. Irani, V. Weerakkody, "Critical Analysis of Big Data challenges and Analytical Methods". *Journal of Business Research (JBR)*, Published by Elsevier Inc. Vol.**70**, pp.**263-286**, **2017**, DOI: https://doi.org/10.1016/j.jbusres.2016.08.001.

[6] S. Rana, M.N.I. Mondal, "A Seasonal and Multilevel Association Based Approach for Market Basket Analysis in Retail Supermarket". *European Journal of Information Technologies and Computer Science*, Vol.**1**, No.**4**, pp.**9-15**, **2021**.

[7] R. Agrawal, G. Johannes, G. Dimitrios, R. Prabhakar, "Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications". *In SIGMOD '98: Proceedings of the 1998 ACM SIGMOD. International Conference on Management of data*. New York, NY, USA. ACM Press, pp.**94-105**, **1998**.

[8] A. Joshi, S. Kale, S. Chandel, D.K. Pal, "Likert scale: Evaluating the Interdependent Effect for Likert Scale Items", *Appl. Sci. Technol*, Vol.**7**, pp.**396-403**, **2015**.

[9] K. Teymoornejad, "Improving Apriori Algorithm with Various Techniques",**2018**. https://www.researchgate.net/publication/329736159_Improving_ Apriori_Algorithm_with_various _techniques.

[10] R. Agrawal, T. Imieliński, A. Swami, "Mining association rules between sets of items in large databases". *Paper presented at the Proceedings of the 1993 ACM SIGMOD. International Conference on Management of data*. **1993.**

[11] B.P. Subedi, "Using Likert type data in social science research:

confusion, issues and challenges". *Int.J. Contemp. Appl. Sci*. Vol.**3**, pp.**36-49, 2016**.

[12] K. Singh, J. Maiti, "Mining Frequent Patterns with Temporal Effect: A Case of Accident Path Analysis". *Proceedings of ICETIT*. Springer, pp.**596-603, 2020**.

[13] Z. Abdullah, A. Gusman, T. Herawan, M. Deris, "MILAR: Mining Indirect Least Association Rule Algorithm". *International Conference on Advanced Data and Information Engineering (ADIE)*, pp.**159-166, 2016**.

[14] C.H. Chee, J. Jaafar, I.A. Aziz, M.H. Hasan, W. Yeoh, "Algorithms for frequent itemset mining: a literature review". *Artificial Intelligence Review*, Vol.**4**, pp.**2603-2621**, **2019**.

[15] A.R. Riszky, M. Sadikin, "Data Mining using Apriori Algorithm for Product Recommendation for Customers", *Journal of computer system technology (JCST)*, Vol.**7**, pp.**103-108, 2019**. https://doi.org/ jtsiskom.7.3.2019.103-108.

[16] D. Patill, R. Deshmukh, D. Kirange, "Adaptive Apriori Algorithm for Frequent Itemset Mining". *Proceedings of the SMART, IEEE Conference ID: 39669 5th, International Conference on System Modeling & Advancement in Research Trends*, College of Computing Sciences & Information Technology, Teerthanker Mahaveer University, Moradabad, India, pp.**7-13, 2016**.

[17] B. Yu, C.F. Wen, H.L. Lo, H.H. Liao, P.C. Wang, "Improvements in patient safety culture: a national Taiwanese survey", *International Journal for Quality in Health Care (IJQHC).* Vol.**16, 2020.**

[18] D. Ginting, H. Mawengkang, S. Efendi, "Modification of Apriori Algorithm focused on Confidence Value to Association Rules". *Nommensen International Conference on Technology and Engineering*. IOP Publishing IOP Conf. Series: Materials Science and Engineering, doi:10.1088/1757-899X/420/1/012125, **2018**.

[19] A. Hodijah, U.T. Setijohatmo, G. Munawar, G.S. Ramadhanty, "Multi-threaded approach in generating frequent itemset of Apriori algorithm based on trie data structure". *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, Vol.**20**, No.**5**, pp.**1034-1045**, **2022**.

[20] R. Likert, "A technique for the measurement of attitudes", *Arch. Psychol*, Vol.**22**, pp.**55, 1932**.

[21] A. Pandey, A. Jain, "Comparative analysis of KNN algorithm using various normalization techniques", *Int.J.Comp. Netw. Inf. Secur*, Vol.**11**, pp.**36-42**, **2017**. DOI:10.5815/ijcnis.2017.11.04.

[22] E. Hikmawati, N.U. Maulidevi, K. Surendro, "Pruning Strategy on Adaptive Rule Model by Sorting Utility Items". *IEEE Access*, Vol.**10**, pp.**91650-91662, 2022**.

[23] M.T. Ogedengbe, S.B. Junaidu F.D.A. Kana, "Itemset Fusion: A New Technique for Itemsets Generation". *ATBU Journal of Science, Technology and Education*, Vol.**12**, No.**1**, pp.**216-230**, **2024**.

[24] D. Kriksciuniene, V. Sakalauskas, "AHP Model for Quality Evaluation of Healthcare System". *In: Damaevičius, R., Mikayt, V. (eds.) ICIST*, Vol.**756**, pp.**129-141**. Springer, Cham. **2017.** DOI:https://doi.org/10.1007/978-3-319-67642-511.

[25] R. Lewandowski, "Evaluating the Interdependent Effect for Likert Scale Items". *Paper presented at the Business Information Systems Workshops: BIS 2019 International Workshops*, Seville, Spain, pp.**26–28**, **2020**.

[26] P. Jalpa, D. Rustom, "A Novel Hybrid Method for Generating Association Rules for Stock Market Data", **2017**.

[27] Z. Kohzadi, A.M. Nickfarjam, L.S. Arani, M. Mahdia, "Extraction frequent patterns in trauma dataset based on automatic generation of minimum support and feature weighting". *BMC medical research methodology*, Vol.**24**, No.**1**, pp.**40**, **2024**.

[28] G.M. Sullivan, A.R. Artino Jr, "Analyzing and interpreting data from Likert-type scales". *Journal of graduate medical education (JGME)*, Vol.**5**, No.**4**, pp.**541-542, 2013**.

[29] K. Sumiran, "An Overview of Data Mining Techniques and their Application in Industrial Engineering". *Asian Journal. Appl. Sci. Technol*, Vol.**2**, pp.**947-953**.

[30] R. Yadav, K Garg, "An Improved Multiple-level Association Rule Mining Algorithm with Boolean Transposed Database". *International Journal of Computer Science and Information Security (IJCSIS)*, Vol.**13**, pp.**9, 2015**.

[31] F. Li, C. Meng, C. Wang, S. Fan, "Equipment Quality Information Mining Method Based on Improved Apriori Algorithm". *Journal of Sensors*, **2023**.

[32] M. Hosseini, O. Akbari, A. Rahmani, A. Shakeri, "Predicting Students' Score in the English Section of the Ba Entrance Exam by Apriori Algorithm". *Research Journal of English Language and Literature (RJELAL)*, Vol.**3**, pp.**306-313, 2015**.

[33] N. Mahajan, J. Namdeo, "Analysis of Student Result Data Using Association Rule Mining". *Research Journal of Science and IT Management, (RJSITM)*, Vol.**4**, No.**6**, pp.**32-37, 2015**.

[34] C.L. Mao, S.L. Zou, J.H. Yin, "Educational Evaluation Based on Apriori-Gen Algorithm". *EURASIA Journal of Mathematics Science and Technology Education*, Vol.**13**, No.**10**, pp.**6555-6564, 2017.** DOI: 10.12973/ejmste /78097.

[35] M.E. Saad, S.N.S. Abdullah M.Z. Murah, "Cyber Romance Scam Victimization Analysis using Routine Activity Theory Versus Apriori Algorithm", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol.**9**, No.**12**, pp.**479 – 485, 2018**.

[36] K. Kakde, P. Gupta, "Performance Evaluation of Student Based on Attendance Using Data Mining Technique", *International Research Journal of Engineering and Technology (IRJET)*, Vol.**5**, No.**5**, pp.**465-467**, **2018**.

[37] N.H. Hanafi, D.J. Juanis, M.H. Samian, S. Mohmad, S.A.M. Shafie, S.N. Shamsuddin, "An Application of Association Rule Mining in Analyzing Courses Affecting the Results of Financial Mathematics Case: UITM Seremban Campus". *Social Sciences*, vol.**9**, pp.**123-133**, **2019**.

[38] F. Asur, D. Sevimli, K. Yazici, "Visual Preferences Assessment of Landscape Character Types Using Data Mining Methods (Apriori Algorithm): The Case of Altınsaç and Inkoy (Van/Turkey)". *Journal of Agricultural Science and Technology (JAST)*, Vol.**22**, pp.**247-260**, **2020**.

[39] D. Bertram, "Likert Scales". http://my.ilstu.edu/~eostewa/497/Likert%20topic-dane-likert.pdf. **2016**.

[40] A.P.U. Siahaan, A. Ikhwan, S. Aryza, "A Novelty of Data Mining for Promoting Education based on FP-Growth Algorithm", **2018**.

[41] S. Jain, S. Shukla, R. Wadhvani, "Dynamic selection of Normalization Techniques using Data Complexity Measures", *Expert System Application*, Vol.**106**, pp.**252-262**, **2018**.

[42] V.G. Manjunatha Guru, V. N. Kamalesh, K. B. Apoorva. "An Ensemble Learning Based Approach for Real-Time Face Mask Detection." *Int. J. Sci. Res. in Computer Science and Engineering (IJSRCE)*, Vol.**12**, No.**3**, **2024**.

[43] S. Pawan Wasnik, S.D. Khamitkar, Parag Bhalchandra, S.N. Lokhande, S. Ajit Adte, "An Observation of Different Algorithmic Technique of Association Rule and Clustering", *International Journal of Scientific Research in Computer Science and Engineering (IJSRCSE)*, Vol.**6**, Issue.**1**, pp.**28-30**, **2018.**

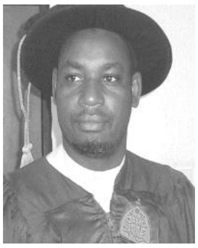
[44] M.T. Ogedengbe, S.B. Junaidu, A.F.D. Kana, "Adaptive Minimum Support Threshold for Association Rule Mining." *Indonesian Journal of Data and Science (IJDS)*, Vol.**5**, No.**2**, **2024**.

[45] S.A. Olaleye, D.C. Ukpabi, O. Olawumi, D.D. Atsa'am, R.O. Agjei, S.S. Oyelere, E.A. Kolog, Association rule mining for job seekers' profiles based on personality traits and Facebook usage *International Journal of Business Information Systems (IJBIS)*, Vol.**40**, No.**3**, pp.**299-326**, **2022**.

[46] L.M. Mohammed, M.T. Ogedengbe, "FP-Growth Algorithm: Mining Association Rules without Candidate Sets Generation.", *Kasu Journal of Computer Science, (KJCS)*, Vol.**1**, No.**2**, pp. **392-411.**

AUTHORS PROFILE

Ogedengbe M.T. earned his B.Sc. (Computer Science) 2009, M.Sc. (Computer Science) 2014, and currently running his Ph.D. (Computer Science) in the department of Computer Science, Ahmadu Bello University, Zaria, Nigeria. He is currently working as a Lecturer 1 in Joseph Sarwuan Tarka University, Makurdi, Nigeria. His main research work focuses on Data Mining, Association Rule Mining, Machine learning. He has 12 years of teaching experience and 10 years of research experience.



Junaidu S.B. Obtained his B.Sc. (Mathematics with Computer Science) from Ahmadu Bello University, Zaria, Nigeria, MSc (Formal Methods in Software Engineering) from Queen Mary and Westfield College University of London and obtained his Ph.D. (Computer Science) from St. Andrews University, Scotland 1998. His area of disciplines includes; Operating Systems, Parallel Computing, Software Engineering, with special skills and expertise in Haskell, Compilers, Online Education, Programming Paradigms and Web Applications Development. He is a Life Member of Association of Computing Machinery (ACM); Member, Institute of Electrical and Electronic Engineers (IEEE); Member, Computer Professionals (Registration Council) of Nigeria; Member, Nigerian Computer Society. He is the Pioneer Head of Department, Department of Computer Science, Ahmadu Bello University Zaria. He has over 51 Research Items on ResearchGate. He has over 25 years of teaching and research experiences. He is currently working as a Professor in the department of Computer Science, and the Dean of student affair in Ahmadu Bello University, Zaria, Nigeria.

Kana A.F.D. earned his B. Sc. (Computer Science) at University of Ilorin, kwara, Nigeria in 2003., his M.Sc. (Computer Science) and Ph.D. (Computer Science) in in University of Ibadan, Oyo, Nigeria, in 2006 and 2015 respectively. His research interests include Knowledge Representation and Reasoning, Machine Leaming, Formal Ontologies and Soft Computing. He is currently working as Professor in Department of Computer Science, Ahmadu Bello University, Zaria, Nigeria. He has many Scopus indexed publications in reputed international journals. He has over 18 years of teaching experience and over 14 years of research experience.

