

# A Fuzzy Classification Model to Solve Customer Relationship Management Problems

Ahmed Abu-Al Dahab<sup>1</sup>  
Business Information Systems  
Helwan University  
Helwan, Egypt

Riham M Haggag<sup>2</sup>  
Business Information Systems  
Helwan University  
Helwan, Egypt

Samir Abu-Al Fotouh<sup>3</sup>  
Accounting Information Systems  
Mansoura University  
Mansoura, Egypt

**ABSTRACT-** The importance of customer relationship management (CRM) to businesses has never been higher. Therefore, companies are forced to adopt new strategies to focus on customers, given the competitive climate in which they operate. In addition, companies have been able to maintain customer data within large databases that contain all information related to customers, thanks to the tremendous technological development witnessed in recent years.

This paper develops a methodology to support CRM to improve the relationship between retail companies and their customers in the retail sector with the aim of retaining existing customers and attracting more new customers, by applying data mining techniques using the genetic algorithm, through which the integrated search can implement the proposed model, because the proposed model It does not need the minimum levels of support and trust required by the user, the proposed algorithm may result in non-redundant, fuzzy multi-level association rules, according to the results of the experiments..

**Keywords:** Customer relationship management (CRM), Fuzzy Association Rule Mining, Multilevel Association rule, Quantitative Data Mining

## 1 INTRODUCTION

Customer Relationship Management (CRM) is a critical component of an organization's information system architecture on which organizations fully rely in order to improve business relationships with customers. Depending on information technology, including the Internet, data warehouses, data mining, etc. After-sales service achieves a lot of the results that are obtained, which enable companies to talk, link and exchange goods and services with their customers through a variety of media, including the Internet, call centers, phone calls, faxes, sales personnel and surveys in order to attract new customers and retain customers. and increase the happiness of customers, and gain the trust and loyalty of customers, which mainly contribute to achieving the greatest return on profitability.

One of the most important factors in the survival of companies is the relationship that binds this company to their customers. It is important to understand that when companies collect a large volume of unmanaged data, such as consumer transactions or sales, decision makers will not benefit much from this type of data. Companies must use modern technologies that help analyze a huge amount of data, process it, and convert it into useful information that can be understood and analyzed, which will give them good management skills and an ideal basic competitiveness in the market [1].

In the information age, data mining is one of the newest and fastest growing sub-solutions in machine learning.

Knowledge discovery from data (information) is a variety of techniques for extracting common patterns from large or high-dimensional data sets. These technologies provide us with accurate information that we may apply in a variety of fields, including business, engineering, and medical sciences. To find new knowledge, different strategies can be used to produce interesting rules known as association rule mining [2].

The Genetic Algorithm (GA) is an indicative model inspired by the theory of natural evolution, and these genetic algorithms are used because they are the most developed and least complex algorithms compared to other algorithms, and for the diversity of their applications and the use of the scientific research method to determine the set of repetitive elements, and this approach is better and easier to use than other genetic algorithms And this heuristic method is used routinely to maintain high-quality responses in order to simplify the research of problems, dealing with them, the ability to address them, and the selection of the best solutions and results. These algorithms are widely used in mining important data to determine association rules. They are also used to find association rules in practical issues such as business databases and fraud detection [3].

Data mining efforts aim to identify association rules at the individual concept level. Finding more comprehensive and meaningful knowledge by processing data and extracting the required information at the same time by exploring the different levels of ideas. Classifications of related elements can be thought of as hierarchical trees, pre-mapped for real-world uses. Inner nodes define classes or concepts that are created from lower-level nodes; [4] the leaf nodes of the hierarchical tree represent the real elements that the transactions are searching for. Figure 1 provides a straightforward illustration of this. There are many reasons to mine multilevel association rules, including (a) the fact that these rules are more logical and user interpretable. (b) We can find answers to unwanted and undesirable rules using the rules of multilevel association. Applications that use spatial data analysis are encouraged [5].

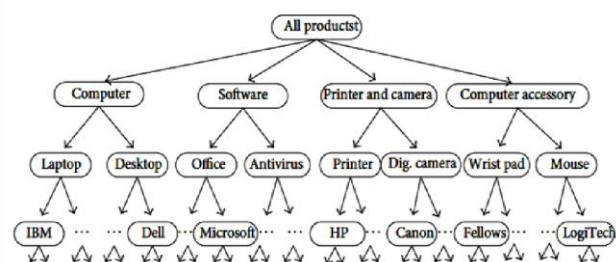


Figure 1. The established taxonomy

### A. Motivation & Rationale

Customer contact and after-sales services have become a major component of corporate strategies due to the increasing competition for retail activities. In the past, companies used to focus more on selling products and services than on looking into the details of the people who made those purchases and neglecting their wants and needs. And because of the abundance of competition, it was necessary for companies to intensify efforts to retain their existing customers, because it was difficult to obtain new customers. In addition, when economic and social conditions were adjusted for, consumer customers were less likely to respond positively to any marketing communications from companies due to their lifestyles. Companies have evolved from product/service focused strategies to customer focused strategies as a result of this environment. Thus, developing lasting relationships with clients has become a major strategic objective. In fact, in order to maintain a positive business relationship with customers, companies that want to stay ahead of the curve must constantly improve service standards [6].

Some companies have invested in building large and huge databases, which enabled them to keep a large amount of data related to customers. Numerous data are collected for each customer, allowing analysis of the consumer's complete purchase history. But the knowledge gathered is rarely used to build enterprise processes such as Customer Relationship Management (CRM). In fact, most companies do not incorporate knowledge into their decision support processes. Information overload and knowledge starvation are recurring problems caused by massive amounts of data. The rate at which analysts are able to process data has lagged behind that data is usable and relevant to the application [6].

Due to this, the research addressed in this thesis aims to develop a method that will assistance with CRM within retail sector, strengthening the link between businesses and their clients. The identification, recruitment, development, and retention of customers are inside of retail topics of this research. To be able to locate well-known objects placed at various levels of abstraction, this essay offers an adaptation for utilizing the Apriori algorithm to mine fuzzy multi-level association rules from big datasets.

### B. Research Problems

Understanding and meeting consumer needs are essential responsibilities for a business to complete in this day of intense competition. As a result, customer relationship management (CRM) now ranks highly among company concerns. How businesses attract new clients and keep hold of existing ones is crucial in today's cutthroat culture. CRM is thought to make the best use of appropriate analytical devices with limited materials to draw the most priceless consumers and heighten their desire to increase purchases. The manager can search and examine a vast amount of data using data mining methods in order to find significant patterns and rules. This approach may also be seen as a means of acquiring knowledge to locate ambiguous association rules to carry out CRM since interactions between the expert system and users can be facilitated by fuzzy knowledge representation..

### C. Problem Statement

Market information the true world typically contains measurable quantities, therefore developing a sophisticated data-mining algorithm capable of working using quantitative data presents a difficulty to researchers that is field [7]. A definition of mining for multilevel association rules.

Follows: Exists a collection of the items  $I = \{i_1, i_2, \dots, i_n\}$  it is a classification tree that succinctly explains the field awareness of the multilayer categorization linkages between things.  $I_1$  is the ancestor of  $i_2$  and  $i_2$  is the descendant of  $i_1$  if there is an edge inform  $i_1$  to  $i_2$ . The database displays just leaf nodes.  $DD$  is a database of transactions, each of which  $TT$  in  $DD$  is a set of items like that  $\subseteq I$ . An identifier is associated with each transaction.  $TT \subseteq DD$ . Let  $P$  shows the collection of positive numbers, and  $I_v$  identifies the set's set identification  $TT \subseteq DD$ . Let  $P$  pair

$\langle x, v \rangle \in I_v$  means the quantitative attribute  $x$ , and the associated value  $v$ .  $I_v = \{ \langle x, l, u \rangle \in I \times p \times p \mid l \leq u \}$ ,  $l$  is the minimum and  $u$  is the maximum of  $p$ . A triple  $\langle x, l, u \rangle \in I_v$  signifies quantitative  $x$  with a value within the range  $[l, u]$ .

Keep in mind that a transaction  $TT$  holds a thing  $x \in I$  if  $x$  is in  $TT$  or  $x$  is a parent to some of the items in  $TT$ .

Moreover, a transaction  $TT$  involves  $X \subseteq I$  if  $TT$  carries everything in  $X$ .

An inference of the form is known as a multilevel association rule  $X \Rightarrow Y$ , where  $X \subseteq I$ ,  $Y \subseteq I$ , and  $X \cap Y = \phi$ . No item in  $Y$  any item in has an ancestor that  $X$ ; that is  $Y \cap \text{ancestors}(x) = \phi$

This is due to a style guideline " $x \Rightarrow \text{ancestors}(x) = \phi$ " is redundant and, with 100% confidence, just somewhat true. Items from any level of  $\Gamma$  can be found in both  $X$  and  $Y$ . [8][9][10].

Quantitative association rule mining still has some restrictions, such as [11]: (1) the design's adoption of a separation of the quantitative attribute prevents it from being usable by all users and attributes. (2) Users, and even specialists, frequently find it challenging to supply certain thresholds, such as the minimal amount of support, curiosity, and confidence. (3) If we use quantitative features, the search space may be very big. Fourth, the algorithm's specified rules may be too numerous to handle.

### LITERATURE REVIEW

Through the book [12] that it is necessary to have a customer relationship management system from the requirement of a comprehensive analysis of the market and the study of consumer needs, product developments, and product life cycles. Information mining is the collection of data on the conduct of product sales over time in order to

analyze different market trends. With this information, a product life cycle can be established and new products with some new improvements can be developed according to market trends and customer desires. Although it is usually a good idea to keep arbitrators global, fuzzy arbitrators operate on a group parameter that can be either general or class-specific. They base their choices on a comprehensive review of the product's sales figure in addition to analyzing other data.

Customer relationship management (CRM) seeks to create a "Learning Relation" with customers in order to help businesses concentrate on their needs, which are the cornerstone of all corporate operations. Businesses put their consumers at the centre of management and operations by monitoring client reactions to specific goods and services. Businesses gain the knowledge necessary to raise the calibre of their goods and services by doing this. In other words, businesses find innovative strategies to keep your current clients, attract new customers, and encourage customers' contributions and loyalty to businesses through regular contact and complete knowledge [13].

Numerous scholars have absorbed on fuzzy multilevel association rules mining in the literature [14–18]. Without specifying the actual minimum support, certain of these solutions suggested multilayer membership functions via systems of ant colonies and genetic algorithms. Setting the functions for each item and then determining the minimal supports are used to increase computing performance. Other projects benefited from the efficiency and versatility that done using OLAP and data mining techniques [8].

Only a few algorithms exist at quantitative multilayer fuzzy association rule mining this time (QMLFRL). Aimed at instance, a higher taxonomic concept requires a minimum amount of support for an item. determined providing the minimal supports for the components related to it in [19], and the total of the minimum supports of the items included in the itemset is used to determine the item minimum support for that itemset. Due to this restriction, descending closure property is preserved, making it easy to extend the original Apriori technique to find fuzzy big item sets.

The authors of [20] provide a brand-new, very innovative genetic-based strategy for choosing criterion values for common item sets. That is method, an advanced coding technique is chosen, both the fitness function is with some assurance. The user-specified minimum support is not necessary for our model. criterion, a thorough search can be accomplished using the genetic algorithm. The experiment's findings show which the suggested approach is able to generate fuzzy multilevel association rules that are non-redundant.

The authors of [5] proposed a new technique for extracting quantitative association rules that may concurrently learn rules and quantize an attribute by using a clustering algorithm. They carried out clustering utilising all qualities simultaneously ahead of time, and they extracted the clusters in the rules from the "association" feature. The authors' technique superior to the traditional Cartesian product type quantization technique in terms of total Rule extraction and quantization accuracy, which was supported by the numerical experiments.

The concept threshold - common item groups which are generated by the evolutionary algorithm - is used in additional pertinent work provided in [3] to develop the quantitative dataset-based rules. In this example, crossover and mutation are used to unify the rule in various ways and can detect the co-occurrence of item sets. Here, comprehensibility, interestingness, and confidence are the three objectives being investigated. As a result, the created laws are developed the principle of multi-objective association. By achieving these goals, the search space for fitness functions is reduced. Finally, distribution-based optimal criteria are developed regarding the numerically valued attribute (A rule's right side displays the distribution of the values for numerical qualities like the mean or variance.)

The advantage of the previous systems is that they contain language terms that make established rules seem much more natural to human specialists; nonetheless, they may produce a significant amount of fascinating association rules. However, since the sparseness of data in three dimensions, Not at all always easy in order to establish efficient association rules (meeting the lowest confidence and assistance) between data points at low (basic) variations in abstraction. Other related issues include: (1) inadequate support for hierarchies that are needed to change over time (2) Real-world application requirements cannot be satisfied by algorithm efficiency; (3) the potential to eliminate the connection between several notion levels; (4) Their approach allowed users to specify various things with differing minimal supports[14][21].

#### Research Contribution

By examining many analytical CRM characteristics in the retail industry, the thesis notion presented in this study makes a contribution to the marketing literature. Additionally, it offers recommendations for businesses on how to utilise utilizing analytical CRM to aid clients in achieving understanding, which promotes the strengthening of the connection with clients. One attempt to effectively find multilevel associations in massive data in the context of CRM is the study's proposed GA-based. multilevel .association. rule. mining. technique

#### II THE. PROPOSED MODEL

This project's main goal was to create a practical Heuristic techniques are used to mine multilevel association rules in huge datasets, with the primary goal of determining the minimal support and minimum confidence levels for each taxonomic level automatically. By utilizing genetic algorithm's capacity to swiftly identify without completing exhaustive searches, numerous solutions can be found simultaneously in a huge multidimensional problem. our proposed a technique can improve mining efficiency while maintaining the desired precision but avoiding the exhaustive list of potential association rules The definitions listed below are those connected to multilevel association rules[8][9][10]:

*Definition 1:* The group of data objects known as an item set  $X \{X_i, X_j\}$ , where  $X_i, X_j \in I$ . Support of an item set  $X$  in a set  $S$ ;  $\sigma(X/S)$  is how many transactions there are (in  $S$ ) it includes  $X$  Compared to the overall volume of transactions in  $S$ . The assurance of  $X \Rightarrow Y$  in  $S_{\sigma}(X \Rightarrow Y/S)$ ,

is a portion of  $\sigma(X \cup Y / S)$  in opposition to  $\sigma(X / S)$ , i.e., the potential for item set  $Y$  which occurs in  $S$  when  $X$  in  $S$  occurs with item set.

**Definition 2:** In set  $S$  at level  $L$ , an item set  $X$  is substantial if  $X$  is supported is not less than the minimal support criterion for it  $\sigma_L'$ . A rule's trustworthiness  $X \Rightarrow Y / S$  this level is high  $L$  if its level of confidence is higher than the minimal confidence criterion for that level of confidence  $\varphi_L'$ .

**Definition 3:** a rule  $X \Rightarrow Y / S$  is robust if  $X \cup Y / S$  at the present level is substantial, and the sureness of  $X \Rightarrow Y / S$  is at the present time high.

**Definition 4:** A fuzzy transaction with the symbol  $T$  is represented by:

$$\overline{T} = \{(x, \mu(x)) \mid \forall x \in I, 0 \leq \mu(x) \leq 1, \mu: I \rightarrow [0, 1], \overline{T} \subseteq T\}$$

$T$  is a broad collection of transactions, and  $\mu(x)$  is level of  $x$  membership.

**Definition 5:** An example of a soft quantitative transaction set is  $T_q'$ . Let  $(F, E)$  a gentle set across the universe  $U$  and  $X \subseteq E, F$  which is the fuzzy power set of  $U \& E$  a set of specifications. A set of attributes  $X$  is seen as backed by a transaction if

$$T_q' = \{ \langle x, l, u \rangle, e \mid \forall \langle x, l, u \rangle \in I \times p \times p \mid l \leq u, e \in E \}$$

The study rules of association focuses mostly on one conceptual level only. A number of some programs that can't find relationships across several abstract planes. It is anticipated that every level of taxonomy's relationships between objects would be discovered in a vast transactional database, any transaction where includes a collection of things and a taxonomy (There is a hierarchy.) on things. Anyone who wishes to research multilevel association rule mining has to be able to afford data at many levels of abstraction and effective multilevel association rule mining techniques. By creating concept taxonomies that go from basic level concepts to higher levels, the first specification can be achieved. The second requirement requires effective multilevel rule mining techniques [9].

The modification of Apriori to multi-level datasets the ML-T2L1 approach [9] [17]. The ML-T2L1 the hierarchical information is encoded into a transaction table that is managed by an algorithm. The dataset's levels are each handled separately. Firstly, Using Apriori, the level one (top of the hierarchy) for enormous It is investigated 1-itemsets. the transaction dataset is then refined by utilising level 1 huge 1-item set lists to remove any items that do not have an ancestor the level 1 huge 1-item sets list, and to trim any transactions that do not have any typical things (thus, when assessed using the level 1 large 1-itemset list, only infrequent items are present.). Level 1 large 2-itemsets finish off the level 1 large 1-itemset list (using the cleaned dataset). Level 1 massive 3-item sets are then implied, until there are no longer any common item sets at level 1, then .Because ML T2L1 requires that only things that are descendants of often recurring level 1 items (essentially,

they have to be descended from huge 1-itemsets at level 1) may remain often occurring their own, the improved transaction table is used to draw conclusions about the sets of level 2 items. Large 1-itemsets are constructed for level 2 which big 2-itemsets are chosen from, followed by big 3-itemsets, etc.the level There are three significant 1-itemsets(from the identical cleaned dataset) once level 2's frequent item sets have all been found, and so forth Either ML T2L1 up till Apriori investigations are conducted on all levels ,or until no significant 1itemsets at a degree of exposure, changes. The following are the main STEPs of the suggested system: [9][10][17][21] [22] [24-26] :

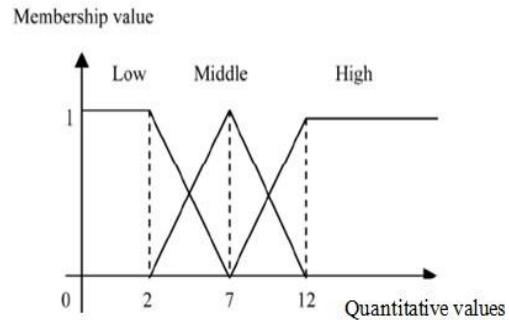


Figure 2 shows how the elements in perform as members.

**Input:** A collection of  $N$  transactional quantitative data  $D$ , a predetermined set  $\Gamma$  containing the original materials  $\{i_1, i_2, \dots, i_n\}$ , a group of membership operations in each case at various levels. In our instance, all membership functions are designed similarly, as seen Figer2; nevertheless, in the  $x$ -axis calculated on the basis of the greater quantitative value assigned to each element in. A minimum support was set using the argument.  $\alpha_k$  and minimal assurance  $\lambda_k$  that a genetic algorithm has gained.

**Output:** collection of fuzzily defined multiple-level association rules with the best minimal support and confidence.

**step1:** Use a combination of integers and the symbol to translate the specified taxonomy. "\*" based on the formula  $C = p * 10 + i$ ,  $I$  is the position number,  $e$  1 of the node at the present time , The letter  $C$  stands for the  $i$ th node's code at the present level. and  $p$  is the  $i$ th node's parent's code at the current level.

**step2:** Analyze the transaction data's term items to determine whether you concur with it's encoded method. when set  $k=1, r=1$  which  $k, 1 \leq k \leq x$  has recently been leveled, The number of tiers is  $x$  a certain taxonomy and  $r$  stands for the quantity of items retained present frequent item sets.

**step3:** the  $k$ -digit entries in a cluster  $D_i$  every transaction, then multiply the item counts in the related sets in  $D_i$ . Symbolise the  $j$ -th total group  $I_j^k$  for  $D_i$  as  $v_{ij}^k$ .

**step4:** We examined a range of membership functions in light of the distinctive characteristics and membership of each type of data item, This results in value transformation  $v_{ij}^k$  each conversation  $D_i$  each group for which an encoding has been made  $I_{ij}^k$  into a fuzzy set  $f_{ij}^k$  (Eq.1) by scheming  $v_{ij}^k$  where depends on the membership function that is

selected  $I_j^k$  is the  $j$ -th item on level  $k$ ,  $v_{ij}^k$  is the quantification of  $I_j^k$  in  $D_i$ ,  $h_j^k$  the quantity of hazy patches for

$I_j^k, R_{jl}^k$  ( $1 \leq l \leq h_j^k$ ) is the  $l$ -th fuzzy the area of  $I_j^k$ ,  $f_{ijl}^k$  is  $v_{ij}^k$  fuzzy membership worth of  $R_{jl}^k$

$$\left( \frac{f_{ij1}^k}{R_{j1}^k} + \frac{f_{ij2}^k}{R_{j2}^k} + \dots + \frac{f_{ijh}^k}{R_{jh}^k} \right) \quad (1)$$

**STEP5:** Create the candidate set by assembling the fuzzy (Language-based) regions with membership values greater unlike zero.  $C_{1k}$ ; Do the scalar cardinality calculations.  $k_{jl}$  of each fuzzy region  $R_{jl}^k$

Transactional data as  $s_{jl}^k = \sum_{i=1}^n f_{ijl}^k$

**STEP6:** Check to see if the value  $S_{k_{jl}}$  every region  $R_{jl}^k$  in  $C_1^k$  exceeds or is equivalent to the threshold  $\alpha_k$  which is achieved by applying the collection of transactions that the genetic algorithm was applied to at this level as the best minimal support for level  $k$  as per  $\Gamma$  (see algorithm 1). If  $R_{jl}^k$  compares to the threshold, insert it in the huge 1-itemset

$L_1^k$  for grade  $k$ . Which is:  $L_1^k = \{R_{jl}^k | S_{k_{jl}} \geq \alpha_k, R_{jl}^k \in C_1^k\}$ . (2)

**STEP7:** Create the applicant set if  $L_1^k$  not empty; otherwise, let  $k = k + 1$  and move on to **STEP3**.  $C_2^k$  from  $L_1^1, L_1^2, \dots, L_1^k$  to catch "crossover level" sets of many items. The created applicant set  $C_2^k$  must meet the subsequent conditions: (1) There must be at least one item in  $L_1^k$  in every 2-itemset in  $C_2^k$ . (2) A 2-item set's regions might not share the same item name. (3) A 2-item set's two-item names might not match the hierarchical connection in the taxonomy. (4) The two huge 1-itemsets that each contain a candidate 2-itemset must have support values that are more than or equal to the minimal support,  $k=2$ , in each case.

**STEP8:** If  $L_1^k$  is empty, then raise  $k$  by one; if not, then set  $r$  to 1; if not, then set  $r$  to  $r+1$ .

(a) If  $r = 2$  make the list of candidates  $C2K$ , where  $C2K$  is the collection of potential goods with 2 items at  $k$  from level  $L11, L12, L13, \dots, L1k$  to discover "level-crossing" of recurring items. In  $C2K$ , each 2-itemset must include at least one item from the  $L1k$  and the subsequent item in the taxonomy shouldn't be its ancestor. Every feasible 2-itemset is made up of  $C2K$ .

(b) If  $r > 2$ , produce the applicant set  $Crk$ , where  $Crk$  is the collection of potential goods with  $r$ -items on level  $k$  from  $Lkr-1$  akin to how it was in the STEPS before.

**STEP9:** For every candidate acquired,  $r$ -itemset  $S$  with items  $(S1, S2, \dots, Sr)$  in  $Crk$ :

a) In each transaction datum, determine the fuzzy value of  $S$ .  $Di$  with the bare minimum operator as  $fis = \min(fis1, fis2, \dots, fisr)$

B) Estimate the scalar cardinality of  $S$  all of the Transactional data as a council  $t_s = \sum_{i=1}^n fis$ .

C) If counts is greater than or equal to the minimum support that has been set  $\alpha_k$  place  $S$  into  $Lkr$ .

**step10:** If  $L_r^k$  if  $K$  and  $r$  are equal to null, then proceed to the next step; otherwise, proceed to step 8 and raise  $r$  by one.

**step11:** If  $k > x$  the next STEP, else set  $r = 1$  go to stop 3

**step12:** produced fuzzy association rules for every common  $r$ -Itemset, including  $S = (S1, S2, \dots, Sr)$ ,  $r > 2$  as follows:

- Catch all the rules  $A \rightarrow B$  where  $A \subset S$ ,  $B \subset S$  and  $A \cap B = \emptyset$ ,  $A \cup B = S$ .

- Calculate the confidence level for each association

rule using by  $\frac{\sum_{i=1}^n \min(f_{iS})}{\sum_{i=1}^n \min(f_{iA})}$ .

**step13:** Select the rules with confidence levels that are greater than the specified confidence threshold.  $\lambda_k$ , where  $\lambda_k$  is the genetically determined minimal degree of confidence for level  $k$ .

**step14:** multi-level datasets with redundant rules should be eliminated. This is Rule  $R_1$  is redundant to rule  $R_2$  if (1) the itemset  $X_1$  is composed of items, at least one of which in  $X_1$  is descended from the things in  $X_2$  and (2) the item set  $X_2$  is totally composed of objects, at least one of which is  $X_2$  is a predecessor of the things in  $X_1$  and (3) the additional non-ancestor things in  $X_2$  are all included in the item set  $X_1$ . The additional state (4) the confidence of  $R_1(C_1)$  equivalent to or less than the degree of  $R_2(C_2)$ .

A. The extraction of parameters using the genetic algorithm

The optimal minimal support and minimum confidence levels are automatically chosen for each taxonomic level using a kind of investigative algorithm known as a genetic algorithm. It investigates a variety of options for dealing with a particular problem. [27]. In order to find a good and better answer, the algorithm creates a "population" of workable solutions to the issue and allows them to "evolve" across many generations. The population, which is a group of solutions (represented by chromosomes), is where the algorithm starts. To create a new population, solutions from one population are chosen and maintained. The fundamental genetic algorithm's structure is as follows: (see Fig. 3).

---

### Procedure genetic algorithm

---

```

begin (1)
  t = 0;
  initialize P(t);
  evaluate P(t);
  While (Not termination-condition) do
  begin (2)
    t = t + 1;
    select P(t) from P(t-1);
    recombine P(t);
    evaluate P(t);
  end (2)
end (1)

```

---

Fig. 3. Structure of the genetic algorithm [19]

- 1) [Start] make arbitrary  $n$  chromosome samples (relevant solutions to the issue)
- 2) [Fitness] Examine the population's fitness (qualifying function  $f(x)$  for each chromosome  $x$ )
- 3) [New population] iteratively develop a new occupant once the new population is complete, the next steps.

A. Selection: Iteratively create a new population through the remaining processes until it is complete.

- B. *Crossover: Crossover the parents in the presence of a crossover probability to create a new generation (children). In the absence of crossover, the offspring accurately reflects the parents.*
  - C. *Mutation: The GA creates a fresh generation of mutations at each spot with a mutation probability (site on the chromosome).*
  - D. *Accepting: a new population that stores different generations.*
- 4) [Replace] control freshly produced population to improve the algorithm's path
  - 5) [Test] If the end condition is satisfied, the program ends and gives the present population its best option.
  - 6) [Loop] go step 2.

The GA maintains a population of results  $p(t)$  during repetition  $t$ .  $r_{1t}, \dots, r_{Nt}$ , where  $r_{it}$  refers to the rule set that is produced at random for each level. The function is used to evaluate each answer.  $E(\bullet)$  and  $E(r_{it})$  is an indicator of how suitable a solution is. A person's fitness value determines whether they have the necessary ability to live and procreate in the next generation. The next iteration is  $(t+1)$ . A new resident is formed based on the procedures (2) and (3) [24].

#### B. Data Encoding

The population in this scenario is composed of several rule collections. The coding method used by the system is called the Michigan approach, and it treats each chromosome as a collection of all applied rules. For each level, With this approach, randomly generated association rules are input. The Michigan technique uses binary coding for its coding, where "1" denotes a knowledge base rule's inclusion in a knowledge base and "0" denotes its exclusion. The main advantage of this methodology is that the entire rule base is codified, therefore it is not necessary to perform a quantitative analysis of all pertinent rules to see whether the strategy works as planned because, Unlike the Pitts method, the genetic algorithm considers every possible rule during the calculation. The large size of the chromosome is a disadvantage. Inversely correlated with the size of the rule base, the size of the chromosome grows exponentially with the number of itemsets. [11][23] [25] .

#### C. Generic Operators

It is common to utilize the genetic operators of reproduction, crossover, and mutation. It is necessary to choice individuals from the target audience to focus on in order to achieve genetic operators. The collection strategy is primarily determined by the population's average level of fitness. The system controls election in in the style of a roulette wheel sample. In this process, Crossover and mutational parents are selected Depending on their fitness i.e., the more fitness function value a candidate has, the greater the likelihood that it would be elected. In order to implement roulette wheel sampling, all applicants' values are first normalised such they have a chance range then a random number between 0 and 1 is estimated by using the random number functionThe candidate is chosen based on a match between both this value and the value adjusted for fitness [25].

Reproduction operators do not transpose an individual as it is chosen; rather, they just replicate it using the present populace in the a fresh populace(The younger generation). Beginning with two chosen people, the crossover operator moves on to the transition spot. (A number between 1 and  $L-1$ , where  $L$  is the length of the strings) is chosen at random. Mutation is the third genetic operator. enables arbitrary shifts in the population configurations it could sporadically produce advantageous outcomes by deviating from a local optimum. Within GA, The only purpose of mutation is to oppose each and every string, Changes a 1 to a 0 and vice versa with a chance of  $pm$  [25].

The method ends working when the state of degradation is reached, or when the disparity the population's top and poorest generating fewer than 0.1% of chromosomes. Additionally, it comes to an end when the user-defined There are now no more generations left. Additionally, the algorithm avoids creating the initial population at random because it can show up in rules that exclude training data in cases when the fitness is very low. Additionally, overfitting the data may result from a population of rules where at least one training instance is guaranteed. The proof demonstrated that non-random methods initialization can result in an improvement in the solution's fitness and can significantly cut down on runtime. [18]. As a result, we developed an initialization approach that selects a practice case to serve as a "seed" for rules creation according to the change in items in each level's item sets [26].

Genetics, generally, operators help to manage population heterogeneity and to prevent early agreement with local optimum [21]. The aim of our investigate intriguing association rules. As a result, It is crucial to consider the fitness function while deciding whether a chromosome is interesting and it does affect how quickly the genetic algorithm converges. The system that is being suggested here investigates two distinct fitness functions. While The connection between the second fitness function's support ( $sup$ ) and sureness ( $conf$ ) characteristics, that are necessary to create A rule for associations, the first fitness function takes into account the same level of assurance as associated association rule as shown in Eq. 4) [8][11][21]. The important variables that balance the importance of the function's backing and assurance are parameters and,  $a+b=1$ . The fitness function's threshold needs to be predefined in order for our GA technique the vast database to mine established association rules; In this example  $a=\beta=0.5$ .

$$f_1: rconf(X \rightarrow Y) = \frac{sup(X \cup Y) - sup(X) \times sup(Y)}{sup(X)(1 - sup(Y))} \quad (3)$$

$$f_2(x \rightarrow y) = \alpha \times sup(x \rightarrow y) + \beta \times conf(x \rightarrow y) \quad (4)$$

By implementing the suggested approach, only the most interesting rules are announced in accordance with the fitness function-defined interestingness measure, as opposed to standard mining models that produce an infinite number of interesting rules. Because GA perform a find and deal globally with attribute cooperation better than greedy rule selection algorithms, they are the preferred method for learning high-level prediction rules [21].

Briefly stated, the proposed evolutionary method for mining quantitative association rules is motivated in

particular by the following: (1) partition of quantitative attribute is not accessible for every attribute and user; (2) users, and even experts, typically feel tedious to define the minimum-support; (3) the search space may be very large when we face quantitative attributes; and (4) the rules passed may be too many to handle [16]. However, mining association rules fall short in terms of benefits; they also have several flaws, most notably a sophisticated algorithm. With the number of things, the number of rules grows exponentially. However, certain sophisticated algorithms are used to tackle this complexity and effectively reduce the search space. The second challenge is choosing intriguing rules from a set of rules, or attaining rules from rules.

The proposed work tackles the second issue, which essentially helps the user scan the rule set, and useful quality controls on the rules are implemented based on genetic algorithms. Usually, managing association rule mining results in a large number of rules being discovered or inferred, confusing the user. More importantly, some of these criteria might not be necessary and produce no new information. Some efforts have been made to address duplicate rules in flat datasets, but redundancy in these datasets needs to be focused on because they can contain a hierarchy/taxonomy or compound idea levels. One of the characteristics of this study is this problem.

### III EXPERIMENTS AND RESULTS

We perform various experiments in this section to evaluate the performance of the proposed strategy and validate the improvements. Experiments are performed on an computer Intel® Core™ i5-2450M CPU @ 2.50 GHz, running on a Windows 7, 64-bit operating system, x64-based processor, and 6 GB of memory. All codes are implemented under MATLAB version 7.8.0.

#### A. Dataset

We have relied on the information contained in [27] because it is considered an important standard to be used for comparison. And this incoming information, which consists of the data of the goods, their quantities, and their elements declared in each purchase container that was marketed, which is known as the market basket data. Each set of data contains more than 1,000 sales receipts for sales in a food warehouse, and each asset contains 7 contracts (10,000 transactions). The pre-determined classification in the first level contains 7 nodes that describe the items that were made in this test, and the second level consists of 14 nodes related to flavor, taste, taste or different types of other stocks and comes in the third level, which consists of 48 nodes that express information about factories and production companies. These transactions contain a database that contains some data such as the name of the product, its shape, and the quantity of the goods that were purchased. However, it is not permissible in any way to use the same element more than once in one transaction

#### B. METHODOLOGY

The methodology proposed by investigating various analytical CRM characteristics in the retail industry, this thesis makes a contribution to the marketing literature. Additionally, it offers recommendations for businesses on how analytical CRM use to facilitate consumers' acquisition of information, which in turn improves the relationship with

customers. This thesis also emphasizes the potential of large-scale database data mining techniques for CRM.

It was compared to the method typically used [27] This study proposes a method for mining multi-level fuzzy quantitative association rules that applies the GA optimization approach. The goal was to (1) enhance multi-level optimum support and confidence employed to obtain interesting rules in order to unearth fresh in-depth data. (2) Removing unnecessary rules that were present in the conventional method. Both approaches consolidate fuzzy boundaries rather than explicit border intervals by inferring large item collections top-down and with increasing depth.

In contrast to the conventional method, which relies on specialists to decide these values manually the suggested approach enables the mining of association rules based on the most advantageous recalculated mining parameters (min sup, min conf). By using GA to determine these characteristics, the suggested system becomes more universal and independent from context. The min-sup and min-conf criteria were chosen at 0.28 and 1.7 for each taxonomic level in the investigations.

In the first experiment, we test the hypothesis that accurate association rules may be given in a bounded amount of time and with a set quantity of first generations. Using our dataset, the initial population sizes vary from 30 to 100 Table 1 displays the results. Our dataset might be used to infer the bulk of strong association rules with a small population. We can determine based on the GA The technique will work similarly to the arbitrary approach insufficiently large population. Despite the fact that comprehensive association rules may be quickly obtained, when the population is too great, the complexity of computing quickly increases. But as we can see, there is a decent alternative. probability that the legitimate association rules will be true with a small population and a short amount of time.

Table 1. CONNECTION BETWEEN THE AMOUNT OF MULTILEVEL ASSOCIATION RULES AND THE NUMBER OF INITIAL GA POPULATION (GENERATION NO. 10 APPLYING MICHIGAN ENCODING WHEN USING F1)

No. of Initial population	30	50	70	100
No. of association rule (Redundant)	3790	610	610	610
No. of association rule (non-redundant)	1702	323	323	323

The fact that the suggested utilizes the Michigan technique for rule coding, whereby a chromosome is a compilation of all applied laws, accounts for part the consistency of the extracted rules' number and a starting population with .50 chromosomes. As a result, the. Fewest starting populations will include certain rule permutations. According to the Michigan method, every chromosome has a detailed representation of the laws.

The suggested system's usage of the Michigan approach for rule coding may be partly credited for the consistency of the recovered rule count with a starting population of 50 chromosomes, where each chromosome is a collection of all applied rules. Therefore, certain rule permutations will be present in the fewest starting populations. Every chromosome has a thorough depiction of the laws according to the Michigan method.

TABLE II : CONTRAST STUDY

Methods	No. Of Non-Redundant Rules	Calculated min-sup	Calculated min-conf	Time (Sec)
Proposed Method with GA	610	L1= 0.96 L2= 0.68 L3= 0.28	L1=1.13 L2=1.45 L3=1.93	550 s
Traditional method without GA [8]	2259	0.29 (L=1 to 3)	1.8 (L=1 to 3)	400 s

Table III COMPARISON STUDY OF THE 2 FITNESS

Fitness Function	levels	Computed Fitness	No. of association rules (non-redundant)
$f_1$ (Eq. 3)	level 1	1.91	610
	level 2	1.98	
	level 3	2.07	
$f_2$ (Eq.4) $\alpha=\beta= 0.5$	level 1	1.07	2259
	level 2	1.27	
	level 3	1.48	

Table .IV EVALUATION OF USING DIFFERENT GA PARAMETERS, THE PROPOSED SYSTEM f1

Parameters ratio	No. of association rules (redundant)	No. of association rules (non-redundant)
Mutation=0.9 crossover= 0.1	760	422
Mutation= 0.8 crossover= 0.2	760	422
Mutation= 0.7 crossover= 0.3	616	347

In line with the aforementioned results, a GA-based technique can more quickly than a conventional technique identify highly relevant association rules (26% more time). Approximately 18% of all rules came by using comparison system. are of higher quality, according the suggested system, which also extracts other intriguing rules. In general, the decision-maker will be hampered by a significant number of extracted rules from market basket analysis. According in relation to a fitness function, which in charge of putting into practice the evaluation that mimics how ideal the answer is: The greater the number the suggested system offers the more exciting laws, the better the solution, and vice versa.

The last set of experiments compared the usefulness of a mined association rule using either the fitness function that takes the associated association rule's relative confidence into account (Eq. 3), or even the fitness function that takes both the support and confidence characteristics into account (Eq. 4). Using the previous GA settings, the experiment is run. The results of Table 3 show that utilizing  $f_1$  further improves the pace at which association rules are mined while reducing the amount of extracted rules by 83%. This experiment taught us that, as demonstrated by the performance of  $f_2$ , For genetic algorithms to succeed, the fitness function is essential. The GA did not benefit from using  $f_2$ , thus we acquire the same amount of extracted rules as we did using the conventional method (about 2259).

Correctly identifying intriguing rules and the calculation in adding to the provision for each item individually, of the union support for the items under each rule are what lead to the enhancement of GA performance with  $f_1$ . Unlike the second function, It encapsulates the typical situation utilized by many of the current mining methods and makes use of each rule's encouragement and assurance (e.g. Apriori).

We will investigate the impact settings for GA factors impact on our system, impact of GA factor adjustments on our system after comparing it to several multi-level quantitative mining algorithms. One parameter setting will be altered at a time while maintaining the default value for another parameter's setting in order to limit the amount of blending factor settings. Table 4 displays how many association rules (non-redundant) our algorithm generated using the dataset given. with mutation rates ranging from 0.7 to 0.9. We can see from the chart that if the mutation rate decreases, there will be fewer extracted rules (18% less rules). This decline is obvious. Because mutagenesis is managed There has been a reduction in the ability of populations to reservation genetic diversity from one generation to the next. Crossover workers are often used in GA too keep an eye on the populace and concentrate on one of the current effective options. Exploration is usually accomplished using mutation operators (exploitation). As a result, the mutation takes every effort to avoid convergence and explore new areas, in contrast to crossover, which strives to concentrate on a certain area of the landscape.

#### IV CONCLUSION

This paper investigated the possibility of using quantitative correlation rules to support CRM managers, where the use of fuzzy theory can improve the decision-making process of CRM managers. The benefit of applying fuzzy association rules is being exploited in a novel approach that manipulates quantitative characteristics by subtly separating attribute values using multi-level association rule mining on huge quantum history ensembles. Fuzzy group theory, multi-level categorization, pre-calculated minimums for each level, and various membership functions for each element are used in the suggested technique. The suggested system gives the user rules based on two interesting criteria, which may be quickly expanded if necessary by altering the trapping function. The findings demonstrated that, in comparison to conventional methods, the proposed method maintains a higher accuracy in the rules of mining associations , Future work involves using CRM is supported by an association rule based on genetic taboos that uses GA to adjust the fuzzy membership function of each element.

#### REFERENCES

- [1] ABHILASHA, Vishal Shrivastava. Customer Relationship Management System Developed using Data Mining and Fuzzy Arbiter. V. Sh. (2019)"
- [2] RAHMAN, Tasnia; KABIR, Mir Md Jahangir; KABIR, Monika. Performance Evaluation of Fuzzy Association Rule Mining Algorithms. In: 2019 4th International Conference on Electrical Information and Communication Technology (EICT). IEEE, 2019. p. 1-4.
- [3] TANWEER, Alam, et al. Genetic Algorithm: Reviews, Implementations, and Applications. International Journal of Engineering Pedagogy, 2020, 10.6.
- [4] E. Mahmoudi, E. Sabetnia, M. Torshiz, M. Jalal, and G. Tabrizi, "multi-level fuzzy association rules mining via determining



- minimum supports and membership functions, " the Second International Conference on Intelligent Systems Modeling, and Simulation, Iran, 2013pp.55-61.
- [5] GARAI, Gautam. Application of Genetic Algorithm in Numerous Scientific Fields. In: Genetic Algorithms. IntechOpen, 2022.
- [6] DEVDA, Pareshkumar; SHAH, Suraj; VASAVADA, Maurvi. ANALYTICAL CRM FOR GOOGLE EDGE-DATA MINING FRAMEWORK WITH REFERENCE TO PHARMACEUTICALS INDUSTRY IN INDIA. International Journal of Management, Public Policy and Research, 2023, 2.1: 40-61.
- [7] Lee, Y. C., Hong, T. P., & Wang, T. C. (2006, October). Mining fuzzy multiple-level association rules under multiple minimum supports. In 2006 IEEE International Conference on Systems, Man and Cybernetics (Vol. 5, pp. 4112-4117). IEEE.
- [8] GARAI, Gautam. Application of Genetic Algorithm in Numerous Scientific Fields. In: Genetic Algorithms. IntechOpen, 2022.
- [9] V. Ramana, M. Rathnamma, and A. Reddy, "Methods for mining cross level association rule in taxonomy data structures" International Journal of Computer Applications, vol. 7, no.3, 2010, pp.28-35.
- [10] S. Saraf, N. Adlakha, and S. Sharma, " Soft set approach for mining quantitative fuzzy association patterns in databases, " International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, issue 11, 2013, pp.359-369.
- [11] X. Yan, C. Zhang, and S. Zhang, "genetic algorithm-based strategy for identifying association rules without specifying actual minimum support," Expert Systems with Applications vol. 36, no. 2, 2009, pp. 3066-3076.
- [12] ABHILASHA, Vishal Shrivastava. Customer Relationship Management System Developed using Data Mining and Fuzzy Arbiter. (2019)". "Volume-2, Issue-6,
- [13] Wang, C. H., & Pang, C. T. (2012). Applying Fuzzy Data Mining for an Application CRM. Bulletin of Networking, Computing, Systems, and Software, 1(1), 46-51.
- [14] P. Gautam, N. Khare, and K. R. Pardasani, " A model for mining multilevel fuzzy association rule in database, " Journal of Computing, vol. 2, issue 1, 2010, pp.58-64.
- [15] WANG, Chen, et al. Mamdani Fuzzy Inference System for Rating the Performance of Sponge City Programme. In: 2022 8th International Conference on Systems and Informatics (ICSAI). IEEE, 2022. p. 1-6.
- [16] S. Prakash, M. Vijayakumar, and R.Parvathi, " A Novel Method of Mining Association Rule with Multilevel Concept Hierarchy", International Journal of Computer Applications, vol. 5, no. 5, 2011, pp.26-29.
- [17] G. Shaw, Y. Xu, and S. Geva, "Eliminating redundant association rules in multi-level datasets, " the 4th P P International Conference on Data Mining, USA, 2008pp.1-8.
- [18] E. Ayetiran, and A. Adeyemo, "A data mining-based response model for target selection in direct marketing," International Journal Information Technology and Computer Science, vol. 4, no. 1, 2012, pp. 9-18.
- [19] Y. C. Lee, T. P. Hong, and T. C. Wang, "Mining fuzzy multiple-level association rules with multiple minimum supports, " Expert Systems with Applications, vol. 34, no. 1, 2008, pp.459-468.
- [20] D. Kanani, and S. Mishra, "An optimize association rule mining using genetic algorithm," International Journal of Computer Applications, vol. 119, issue 14, 2015, pp.11-15.
- [21] Y. Wan, Y. Liang, and L. Ding, "mining multilevel association rules from primitive frequent itemsets," Journal of Macau University of Science and Technology, vol.3, issue 1, 2009,pp.10-19.
- [22] GAO, Qiang; ZHANG, Feng-Li; WANG, Rui-Jin. Mining Frequent Sets Using Fuzzy Multiple-Level Association Rules. Journal of Electronic Science and Technology, 2018, 16.2: 145-152.
- [23] S. Manish, A. Agrawal, and A. Lad, "Optimization of association rule mining using improved genetic algorithms," the International Conference on Systems, Man and Cybernetics, vol. 4, USA, 2004, pp.3725-3729.
- [24] M. Kayaa, R. Alhajj, "Genetic algorithm based framework for mining fuzzy association rules, " Fuzzy Sets and Systems, vol. 152, 2005, pp. 587-601.
- [25] S. Tiwari, M. K. Rao" Optimization in association rule mining using distance weight vector and genetic algorithm , "International Journal of Advanced Technology & Engineering Research, vol. 4, Issue 1, 2014, pp.79-84.
- [26] P. Wakabi-Waiswa, and V. Baryamureeba, "Mining high quality association rules using genetic algorithms", In Proceedings of the twenty second Midwest Artificial Intelligence and Cognitive Science Conference, USA, 2009, pp. 73-78.
- [27] GAO, Qiang; ZHANG, Feng-Li; WANG, Rui-Jin. Mining Frequent Sets Using Fuzzy Multiple-Level Association Rules. Journal of Electronic Science and Technology, 2018, 16.2: 145-152.