J. Electrical Systems 20-10s (2024): 5812-5834

[1] Rajkamal Sarma[1], Pankaj Kumar Deva Sarma[2*] Nayanjyoti Mazumdar[3]

Study of Algorithms for Mining Fuzzy Association Rules and Applications

Abstract: **-** Knowledge discovery in the form of rules has become a useful and meaningful practice in Data Mining. In the last three decades, Association Rule Mining has been considered one of the primary research activities. Introducing fuzzy mathematical theories provides another dimension to make association rules more user-friendly and expressive. Membership degrees and Linguistic Terms play a significant role in the fuzzification process. The use of Fuzzy set concepts in rule generation, however, needs more attention from the researchers in different application domains. This paper is based on the study and findings of some important Fuzzy Association Rule Mining (FARM) algorithms and their applications. Beginning with the overview of Fuzzy Set theory and FARM algorithms, this study gives an analysis and account of the intricacies of algorithm development and applications-based activities in the last three decades. Out of many algorithms, some important FARM algorithms are undertaken for the study and explained with examples. The examples are prepared from both real and synthetic data. Based on the study, some key features of leading algorithms are observed and highlighted. Different research issues and challenges related to FARM are found out and discussed for further research.

Keywords: Fuzzy Association Rule, Data Mining, Membership Degree, Linguistic Term.

1. INTRODUCTION

The large-scale growth of data in various applications has created challenging issues in the discovery of valid and effective association rules through the processes of data mining. Knowledge Discovery in Database (KDD) [1] plays a vital part in the transformation of meaningful data. Three key steps are implemented in the KDD process for the generation of association rules. Firstly, pre-processing of data, secondly, mining of data, and finally, post-processing of data. Among these steps of KDD, data mining focuses on the extraction of patterns that are understandable, valid, novel, and useful. Techniques of data mining include discovering association rules, classification, clustering, regression, summarization, image retrieval, functional dependencies, rule extraction, etc. [2]. Since its introduction by Agrawal and his co-researcher in [3], the techniques of ARM have drawn considerable attention from researchers and in this connection, different algorithms are proposed. The algorithms are about finding interesting relationships among different attributes present in transactions of large databases and representing these in the form of association rules. Depending on its nature and working process the proposed association rule can be classified into different categories like Boolean, Generalised, Quantitative, Temporal, Fuzzy and so on. Several publications related to FARM have taken place and various algorithms to generate rules have been proposed in the last three decades. These along with developments taking place till date are considered as part of the survey work in this paper.

The motivation behind surveying different fuzzy ARM algorithms and applications is to explore the effectiveness of these approaches in handling uncertain data, understand its features, and analyze different FARM algorithms and applications across different domains. The basic objective of such a survey is to provide an overview of some leading algorithms and to identify the strengths and weakness, issues, and complexity of those algorithms. In addition to this, another reason for this survey is to find the possible research gap in the execution process in different domains and explore the possibilities of further research. Overall, this paper is an attempt to study some important algorithms within the realm of Fuzzy association Rule Mining and identifying key features with application in different domains.

The organization of the paper is summarised as follows: section 2 describes the overview of related works of both classical and fuzzy association rule mining techniques along with their application in various domains. The fundamental concept of Fuzzy Association Rule Mining about fuzzy set theory is presented in section 3. Some common approaches, followed by selected leading algorithms are mentioned in section 4. In subsections 4.1 to

[1] [1,2*,3]Department of Computer Science, Assam University, Silchar, Assam, [1]E-mail: rajkamal_sarma@rediffmail.com,

[2]E-mail: pankajgr@rediffmail.com, [3]E-mail: nayan.mazumdar@gmail.com

4.5, FARM algorithms are explained with examples. In section 5, a summary of the algorithms selected for this study is discussed and the key features of those algorithms are highlighted. In section 6, some research issues are identified based on the survey as the research gap. Finally, in section 7, the summary of the entire work is concluded.

2. Related Works

With the rapid increase in the growth of data in every application, converting data into proper knowledge has already become a challenging issue. Rule mining is a key function in data mining practices where rules can be mined or generated from data for the discovery of the relationship among the attributes of a transaction dataset.

Hence, association rule generation has significance in data mining and its different research activities where interesting relationships among various attributes are explained **[**3**]**. In Market-Basket analysis, by applying techniques of association rule mining, the buying habits or patterns of customers are analyzed. For this, associations between items bought by customers are found in the sales transactions. To find patterns using association rule mining technique it needs to identify rules in the form of A→B in transactional data. Such a rule means that if an item contains an attribute A, then it also has a tenancy to contain attribute B, for example, bread → milk. However, to generate association rules, some important quality measures viz. support and confidence are required. Depending on data representation, association rules can be classified into different types like Boolean, generalized, quantitative, etc. All these types of association rules have some limitations to discover nontrivial knowledge. A database containing values between 0 and 1, can be extended to the classical mining algorithm using Fuzzy set theory. Thus, imprecise terms and relations can be effectively modeled in communication and understanding by utilizing Fuzzy sets, allowing for optimal representation **[**4**].** Hence, the simplicity of knowledge representation is a key factor that has led to the recognition of fuzzy-based techniques an important component of data mining system**s [**5**]**. Another algorithm called IFWARM proposed by T. Yang and et. al, is an improved version of the standard Fuzzy weighted association rule mining algorithm and provide more stability and effectiveness is found in IFWARM using the Weighted DCP [6]. G.Sumathi and et. al [7] proposed the IFWIAR algorithm which was executed based on a recommendation model. With meaningful queries collected from the web and then pre-processed and analyzed those queries, recommendations can be made. N.Allahverdi[8] applied a fuzzy approach in medicine to improve the quality of the diagnostic and services. Fuzzy Expert system (FES) has been applied in diagnosis using different linguistic terms to classify different ranges of parameters. An attribute-oriented approach was investigated to KDD by J.Han and et. al [9] and applied in generalization to reduce the computational complexity of the database learning process.

Fuzzy Association Rule Mining algorithms have been applied in various real-life datasets in recent times as it is simple to understand the rules and relevant in practice. C.T. Dhanya and D. Nagesh Kumar [10] applied an FP-Tree-based algorithm for the prediction of the monsoon rainfall in India. M.Sulainman khan et al. [11] proposed an efficient fuzzy algorithm for healthy association rule mining and applied it to nutrient-based data to find associations by converting a transactional database into a database having average RDA (Required Daily Allowance) of nutrient values per item. CFARM algorithm is very useful in composite dataset like food nutrient [12] to find associations among different nutrients. R. Sharma et al. applied the CFARM algorithm among nutrients of different food recipes [13]. Guoqing Chen et al. [14] focused on the uncertainty in fuzzy taxonomies and implication-based fuzzy association rules with functional dependencies for controlling data similarity and unnecessary data. Ashish Mangalampalli and his co-author Vikram Pudi [15] presented the FAR-miner algorithm where they mentioned that the FAR-miner algorithm is faster for large fuzzy datasets. Many algorithms are used to mine association rules in relational and transactional data separately. K.C. and et. al [16] presented an approach called Fuzzy Miner to mine such rules in both kinds of data and suggested that Fuzzy Miner can be executed to a real-life database that contains both relational and transactional data. Bin Pei et al. [17] designed the FARP Algorithm which is suitable for finding frequent fuzzy-probabilistic quantitative item sets from all kinds of databases like quantitative, categorical, or Boolean. Peng Yan et al. [18] introduced both positive and negative fuzzy association rules. The use of the negative rule reduced the database scan and made the algorithm efficient. Ruchi Bhargona et al. [19] considered three basic limitations found in most of the algorithms and designed an improved model of the algorithm which resolved that limitation by generating both positive and negative association rules based on weighted support and confidence and accepting not only Boolean dataset but also considering quantitative data. In a survey paper, Bodrunnessa Badhon et al. [20] go through of a review of recent multi-objective evolutionary algorithms by listing a comparative analysis of encoding, objectives, generic operators, and datasets on categorical, quantitative, and fuzzy rule-type variables. N. Sundaravalli and A. Geetha [21] used different data mining techniques like clustering and classification to analyze the prediction of rainfall and crop production as well. Anna L. Buczak et al. [22] proposed a modified algorithm and extended a model for predicting malaria disease. The researchers demonstrated the approach to use for the prediction of such different diseases. The fuzzy association rule mining concept also can be applied in marketing. Carmen Kar Hang Lee et al. [23] applied this idea in the case of a study conducted in a Hongkong based fashion company where rules were generated on a real dataset. The fuzzy-based approach was found to be very useful for knowing customer satisfaction and thus helps a company to make business policies to attract customers. Guney Gurset in his review article "Health care, uncertainty, and fuzzy logic" [24] mentioned the descriptive study for examining different approaches of fuzzy logic in healthcare. In the paper, "User Interacting Navigation Pattern Discovery and fuzzy correlation-based Rule Mining" [25], the author Uma Maheswami and Dr. R. Gunsundarri proposed a fuzzy correlation algorithm to avoid misleading rules generated by traditional approach from web log file showing the minimal usage of space taking cost and time as a criterion. In the paper [26], Xinhui kang et al. apply two methods in their study to improve the satisfaction of customers on product form by using the Fuzzy Delphi Method (FDM) for filtration of important items and fuzzy ARM method to find the fuzzy weighted rule. Akash Saxena and Vikram Rajpoot [27] performed a comparative analysis of different Association rule mining algorithms. In their survey, different algorithms like AIS, SETM, Apriori TID, Apriori hybrid, FP Growth are studied and comparison is done among these algorithms to focus on some factor like accuracy, speed, support of database etc. Tao Pan, in his paper [28] applied an improved Apriori algorithm by using the divide and rule principle. In the paper, physical educational data was analysed with both original and improved Apriori algorithm to find correlations of different indices of physical fitness experimented in college students of China. Fuzzy Association rule mining with integrated process mining and fuzzy multi-attribute decision are applied to identify the anomalies in business processes mentioned by Riyanarto Sarno et al. [29**]** In a comprehensive survey [30], by Manoj Kr. Gupta and Pravin Chandra have formulated some challenges for data mining researches. In the paper "Performance evaluation of Fuzzy Association Rule Mining Algorithms" [31], Tasnia Rahman and et.al presented a comparison analysis of classical algorithms with two fuzzy-based algorithms on two datasets. According to the authors Genetic Fuzzy Apriori Association Rule Mining algorithm is more efficient among those three algorithms. G.T. S Ho et al. [32] proposed method that enhance the investors to study the associations among different parameters and further can help in applying the knowledge in their decision support system. The case study was done comparing the Hang Sen Index of Hong Kong with other economic indices. It can also help the investors to find hidden pattern and rules in decision support system. S. Vinodh and et.al [33] considered different quantitative and qualitative agility attributes as a case study with their method to find rules from database for agility evaluation. It also helps to make flexible decisions based on different attributes. The application of the fuzzy technique in finding the associations of various socio-demographic conditions is very useful for researchers to generate knowledge discovery in the form of rules. In the recent Covid-19 pandemic situation, these socio-demographic conditions have some effect on the spreading of COVID-19. S. Chatterjee et al. [34]. collected such socio-demographic data from 13 countries and experimented with the FARM technique to find different relations among different socio-demographic data. However, more experiment work to be done, according to them. In a study of people’s attitudes through social networking regarding anti-vaccination towards COVID-19, the FARM technique was applied by N.Eliguzel[35]. XUE Yue-Jn et al. applied the weighted fuzzy association rules [36] method for the improvement of accuracy level by experimenting with the soil quality. Thus, finding the fuzzy rules for soil quality assessment, rules are mind, and redundant rules or insignificant rules are eliminated. Another algorithm proposed by J.Pillai et.al.[37] called Fuzzy High Utility Rare Itemset(FHURI) is an extension of HRUI algorithm proposed by the same authors. Practically, this algorithm is useful for minimizing the cost of high-utility rare items, assessing of quality of goods and services, and identifying the itemset that are profitable. It is very handy in the prediction of the e-commerce world. Onur Dogun and et.al [38], proposed FARM model and demonstrated how it can be more effective and more informative in E-Commerce compared to traditional ARM. Considering E-Commerce as a trending mode of business with advancement in technology, O. Deogan [39] proposed a new method called P-FARM Profit- Supported Association Rule Mining with Fuzzy theory. The technique generates association rules based on the most profitable items in frequent items. Deepesh Kumar Srivastava and et.al [40] applied the fuzzy technique to process 35 years of economic data for India collected from the World Bank website to mine important associations among some economic development indicators of India on the quantitative data between 1979 to 2013.Peng Yan and et.al [41] introduced negative association rules along with positive association rules. The negative association rule was quite useful to reduce the database scan and make the algorithm effective. A novel method called LOGIC was proposed by X.Dong et.al [42] to control redundancy of positive negative association rules. According to the experiment results 81.6% of redundant rules were pruned out by this method. C. Chen et al [43] proposed a mining algorithm on transactional data which was applied to the customer's preferences in online shopping. Some linguistic terms were used to classify the preferences to convert them into fuzzy values. Thus, using fuzzy linguistic assessment helps decision-makers to get interesting knowledge. P.Matapukar and S. Srivastava[44] studied two different datasets and compared different parameters with different algorithms. S. Princy et al [45] performed a comparative analysis between triangular and trapezoidal membership functions to fuzzified datasets. EsralAkguil and et.al [46], in their case study of Cradle design, applied the concept of fuzzy linguistic summarization method to discover fuzzy association rules using the Apriori algorithm for Kansei Engineering, which deals with customer's psychological feelings and needs. Z. Zhao et al. [47] proposed an improved Apriori algorithm considering some data mining tools like dataset size, technical conditions, etc. The proposed algorithm has an advantage regarding search speed of finding reduced frequent itemset. A novel fuzzy methodology-based FARM method for biological knowledge extraction was proposed by F.J. Lopez et al [48] and applied the methodology over a Yeast genome dataset. In most FARM algorithms, it is assumed that a fuzzy set is given. M.Kaya et al [49] proposed an automated method for autonomous mining of both fuzzy sets and fuzzy association rules. To perform such a method, the clustering algorithm used first before membership function. A algorithm called CBFAR (Cluster-based Fuzzy Association Rules) approach was proposed by H.P.Chiu and et al [50] to address some issues found in the transaction and online browsing data in the field of E-Commerce. In the health sector, to assess the risk factor correlated with the disease a technique called PB-FARM (Profile Based Fuzzy Association Rule Mining was proposed by A. Yavari et al [51]. The article [52] examines the role of descriptive data mining in enhancing the maintenance and reliability of physical systems. It reviews the use of association rules and their applications in industrial settings, while identifying key research gaps. The study [53] reveals a correlation between classification accuracy and certainty in fuzzy associative classification, with greater certainty leading to better accuracy. This highlights the relevance of applying fuzzy logic in ARM. In the paper [54], authors reveal a correlation between classification accuracy and certainty in fuzzy associative classification, with greater certainty resulting in better accuracy, emphasizing the importance of fuzzy logic in ARM. The article [55] explores various factors influencing academic success using student data from diverse backgrounds and assesses the accuracy of the proposed mode. The paper [56] discusses the FARM algorithm and its use in generating meaningful rules to address sharp boundary issues in classical ARM techniques. The fuzzification method converts quantitative datasets into fuzzy and binary categories, thereby developing and analyzing rules.

3. Fuzzy Association Rule Mining

The use of the Fuzzy concept in Mining Association Rules is become an improvised approach for the discovery of association rules developed in recent years. This approach is developed to address the issue of mining quantitative data frequently present in databases. Some algorithms are already been proposed by different researchers to extract rules from quantitative data. In this approach, attributes of the dataset are classified into different categories based on an assumed range of values, which leads to a sharp boundary problem. Nearby values of these crisp sets are generally ignored or overemphasized. The academic performance of students is traditionally evaluated based on a percentage of marks, typically falling within the range of 60% to 80%. However, the conventional method of categorizing students based on a sharp boundary approach can be limiting. For instance, a student scoring 61% and another scoring 79% would be considered to be of the same quality, whereas a student scoring 59% would not be categorized similarly. This issue is known as the "Sharp boundary problem" in the context of crisp sets. To address this, fuzzy concepts are employed, allowing for the representation of values between 0 and 1, thereby providing a more nuanced approach to grading and membership in multiple sets. This approach is also beneficial in dealing with categorical data, where items are assigned to specific categories, such as distinguishing between papayas and cucumbers. Furthermore, the use of proper linguistic terms can help in addressing the binary association rule, making it more comprehensible to humans, especially when considering Boolean values for attributes. Most of the rule mining algorithms focus on positive associations among transactional databases and generate the rules according to those positive associations between Itemset. However, F-PNWAR mining algorithm [57] emphasized both positive as well as negative association rule mining, which is a new addition to rule mining techniques. E-FWARM is another modified and enhanced fuzzy-based approach that is very useful in weighted association rule mining [58]. The highlight of this algorithm is to assignment of proper weight of the itemset based on significance which leads to more meaningful rule generation from the database. Interestingness measure framework is another issue of the traditional FARM approach to generate rules. Support-confidence framework is very common and sometimes may not be so effective. In this regard, Fuguang Bao and et.al [59] analyzed the merits and demerits of traditional measures used in classical ARM and proposed some new parameters that are found to be better. Considering these above-mentioned situations bubbled in quantitative, categorical, and binary association rules, the Fuzzy technique or fuzzy concept may be more effective. Hence, FARM is termed as a mining technique that is based on fuzzy set theory and fuzzy logic to generate some proper useful rules.

Fuzzification is a method that converts crisp values into fuzzy ones, addressing the boundary problem between different attributes in Crisp set theory. It uses the membership function, $\mu(x)$, to map the crisp set of intervals between [0,1] into different linguistic terms with varying participation degrees. For instance, the performance of students in a class can be categorized as poor, medium, or good based on the percentage of marks obtained. The percentage intervals are fixed for each category, such as (30%-45%) for poor, (45%-60%) for medium, and (60% and above) for good. Now, if a student scores 59.5%, using a crisp set approach, the student would be classified as medium, even though they are just 0.5% away from the “good” category. This sharp boundary in crisp sets can sometimes lead to inaccurate or unfair classifications. Fuzzy sets address this issue by assigning grades between 0 and 1 based on a membership function, allowing for a smoother and more accurate representation of the data The fuzzy set theory was proposed by L.A. Zadeh [60] to give an approximate, but effective way for describing the characteristics of a complex system to perform precise mathematical analysis [61]. Fuzzy set theory was applied by G.Pasi [62] in different information retrieval to define flexible systems. The use of the Fuzzy set technique in different E-Commerce applications and development in various aspects was explained by J.Lu [63]. Membership Functions is considered an effective method in fuzzification of crisp value and representing them in different intervals. Modeling fuzzy systems using conventional mathematical techniques is very complex. Fuzzy processes are vaguely defined. Description of such processes have some uncertainty. Fuzziness addresses those objects and processes, which cannot be defined distinctly. In crisp sets also there is some fuzziness regarding membership, boundaries, or categories of sets, that are solved by using fuzzy sets. The complexity of fuzziness can be simplified by making a satisfactory trade-off or compromise between the information available and the amount of uncertainty. Thus, fuzzification provides more flexibility and simplicity connecting human reasoning to intelligent systems.

4. Leading FARM Algorithms

Discovering proper association among attributes in transactional data acquired the prime attention of the researchers since the formulation of the problem, often called "the market-basket problem" [3][64]. Several algorithms are designed to discover the association rules by applying different strategies. Algorithms like Apriori, Apriori TID [65], etc. were developed to improve the previous approaches. However, the incorporation of fuzzy sets has brought about a substantial transformation in rule mining techniques and their corresponding algorithms. The introduction of fuzzy set theory in association rule mining algorithms mostly involves the following steps:

i) Pre-processing of raw data as sample dataset and presented in the form of transaction dataset

ii) Conversion of the transaction dataset into a fuzzy dataset using the membership function.

iii) Classification of fuzzified value into linguistic terms.

iv) Frequent itemset generation based on the measures of interestingness.

vi) Generating Fuzzy Association Rules with natural and meaningful representation.

Some leading algorithms which are proposed by different researchers are listed as follows:

Table.1 . List of Leading Algorithms.

SL No.	Name of the Algorithms
1	Fuzzy Automatic Pattern Analysis and Classification System(F-APACS) [66]
2	Fuzzy Transaction Data Mining Algorithm (FTDA) [67]
3	Fuzzy Quantitative Association Rule Mining (FQARM) [68]
4	Composite Fuzzy Association Rule Mining (CFARM) [69]
5	Fuzzy Weighted Association Rule Mining (FWARM) [70]

4.1 F-APACS

The algorithm known as F-APACS [66] was introduced to extract Fuzzy Association Rules from the quantitative database. This algorithm has some uniqueness, especially in the use of linguistic terms and adjusted difference concepts. Fuzzy technique is regarded as vital part of data mining as it provides simplicity to human knowledge representation [5]. The use of proper linguistic terms can make the rules more understandable to the human. To find association among different attributes, some user-defined interestingness measures like support, and confidence is commonly used in generating association rules. In this process minimum value of support count or confidence may leads to some unnecessary noisy rules and their irrelevant relationship as well. It is a big challenge for the researcher to decide the proper support and confidence value. To address this issue F-APACS brought some new measures like adjusted difference analysis and weight of evidence [71]-[73] which had been used to find associations among its attributes. The algorithm allows users to extract rules based on Positive association and Negative association. This is regarded as a key feature of the FAPACS algorithm. Because most of the algorithm discover only positive association rule. Positive association rules simply mean that for any attribute value of a record there is another attribute value whereas in the case of a negative association rule, there is no attribute value against another attribute. The attribute value may be represented as a linguistic term. The f-APACS approach is useful for mining linguistic data. This data is easy to understand. The concept of linguistic term is defined based on fuzzy set theory and because of this, FAPACS is considered as one kind of FARM approach. Here, to explain the algorithm, we have assumed an example of a college library as a sample dataset. In a library, there is a record-keeping database of readers, classification of reading materials like books, magazines, newspapers, journals, etc, reading hours of different readers, and duration of study. Depending on the reading behavior or pattern of readers, there may be some rules that may help the librarian to run the library smoothly. Among different attributes, we consider some attributes to know how F_APACS algorithms work and provide some rules. The attributes are Type_ of_ readers, Type_ of_ reading, Time-of_ reading, and Duration of-reading. Here, the attributes Type_ of_ readers and Type_ of_ readings are categorical attributes whereas the remaining two attributes, time_ of_ reading and Duration_ of_ reading, both are quantitative. The domain of type_ of_ readers is {Student, Teachers, Officials, Guest} and the domain of Type_ of_ readings is {Textbook, Reference Book, Journal, Magazine, Novel, Newspaper}. Let consider the other quantitative attributes time of reading and Duration_ of reading. The total time duration (9 AM to 3 PM) is divided into some shifts as shown in the fig1.

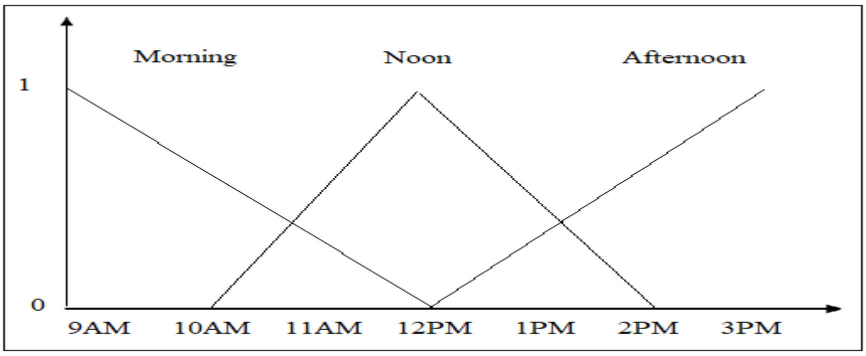

Figure1. Membership Function used for Time_ of_ reading.

The Figure 1. Represents that the attribute “Time_of_reading" is converted as per the triangular membership function into three categories Morning, Noon and Afternoon. The figure also indicates that the attribute "time_of_reading" is gradually changing with membership degree 0 to 1. Consider 12 noon as strictly high, i.e., "1" and it is categorized as noon. Similarly, Morning and Afternoon shifts are also classified. The attribute "Duration_ of_ reading”, is also classified into linguistic terms like short, moderate, long w.r.t. membership degree in between 0 to 1, as shown in the figure.2.

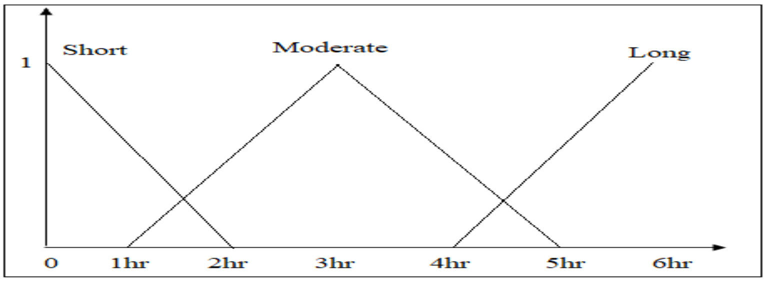

Figure2. Membership Function used for Duration_ of_ reading

Here, the triangular membership function is used to classify the Duration _of _reading in hours. Now, by, applying the F_APACS algorithm, some associations can be shown:

Type_ of_ readers= “students” ⇨Type_of_reading= “Textbook.”

Type_ of_ readers= “Teachers”⇨ Type_ of_ reading = “Journal.”

Duration_ of_ reading= “short”⇨Type_of_readers= “Office_ staff.”

Type_ of_ readers= “student” ⇨ Time_ of_ reading= “Morning.”

4.2 FTDA

Fuzzy Transaction based Data Mining Algorithm (FTDA) is also an important and simple technique proposed by Hong et al [67] [74][75]. The algorithm is most often applied to quantitative data. Here, the Algorithm (FTDA) is applied to rainfall data measured in millimeters. An experimental workout of the FTDA algorithm is explained to generate some useful association rules. The data contains the average rainfall during monsoon seasons in ten years span from 2011 to 2020 collected from BNCA station under Assam Agriculture University, Jorhat (Assam). The sample dataset includes 10 transactions as shown in Table 1. Here, it consists of rainfall data of four months i.e. June, July, August, and September. In the mining process, each month is considered as an attribute. In this example, each month of the monsoon season is considered as an attribute and the amount of rainfall are decomposed into three categories: low, medium, and high. So, according to the predefined membership function shown in the fig.3. these fuzzy membership degrees are calculated for each attribute.

Table 2. Sample of Rainfall Data measured in mm

TID	June	July	August	Sept
1	6	15	15	8
2	15	10	11	10
3	5	17	7	6
4	11	13	11	13
5	16	13	18	9
6	10	12	9	9
7	16	10	11	10
8	10	13	11	7

9	6	10	5	10
10	14	9	6	8

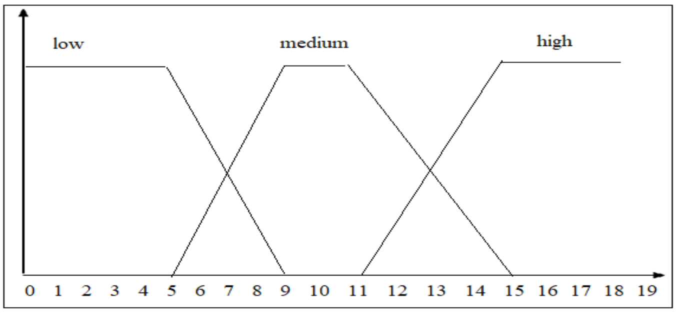

Figure3. Membership Function used for rainfall data

Step1: At first the values in each transaction are converted into fuzzy sets. For Example, value 6 is transformed into fuzzy value June_low=0.8 and June_medium=0.2

Step 2: Here, the total count value for each category of the dataset is calculated by aggregating the values found in each transaction for a particular category, and the same process is done for another category too. For example, the scalar cardinality of the region of June_low can be calculated as (0.8+0.0+...+0.0) =2.6

Step 3: Thus, the highest count is calculated from three possible regions for each attribute. In the case of June, the count is 2.6, 3.4, and 4.0 for low, medium, and high respectively. Since the highest value is found against "June_high" among all the attributes for June, hence "June_high" will be considered for the mining process. The same step is followed for other attributes too and thus Four regions are selected. These are given in the following Table 2 with their support count.

Step 4: Now the support counts of different regions are checked with the given threshold which is 3.5 in this example and since all the selected regions have more support count than 3.5, therefore, all the five regions will remain the same for the next step of the mining process.

Table. 3. Items having the highest cardinality with support count

Sl No	**Item set**	**Support**
1	June_high	4.0
2	July_medium	5.6
3	August_medium	4.6
4	Sept_medium	6.8

Step 5: let set r=1.

Step 6: Now, the Candidate sets Cr+1 from Lr generated: Initially, C2 is generated from L1 as follows:

(June_high, July_medium), (June_high, August_medium), (June_high, Sept_medium), (July_medium, August_medium), (July_medium, Sept_medium), (August_medium, Sept_medium)

Table 4. The Itemset and Fuzzy Count L_2

Sl No	**Itemsets**	**Count**
1	(June_high, July_medium)	3.4
2	(June_high, August_medium)	2.0

3	(June_high, Sept_medium)	3.6
4	(July_medium, August_medium)	3.2
5	(July_medium, Sept_medium)	5.4
6	(August_medium, Sept_medium)	3.6

Step 7: From the previous step some newly formed candidate sets are found where the following sub-steps are executed:

a) Using intersection, the fuzzy membership value is calculated for each transactional data.

(b) The counts are compared with a pre-defined minimum support value of 3.5. This results in three Itemset (June_high, Sept_medium), (July_medium,Sept_medium) and (August_medium,Sept_medium) as L2 as in table 5.

Table. 5. The Fuzzy Count of Itemset in C2

Sl No	Itemset	Count
1	(June_high, Sept_medium)	3.6
2	(July_medium, Sept_medium)	5.4
3	(August_medium, Sept_medium)	3.6

Step 8: The next step depends on whether a large frequent item is null or not. If Lr+1 is null then we must move next step; else, r is set as r+1, and step 6-8 need to be repeated. Here, as L2 is null we cannot go further.

Step 9: Some associations for each large item are found using the following sub-steps:

(a) Among the possible associations, the following rules can be found:

If June=high, then Sept=medium.

If Sept=medium, then June=high.

If July=medium, then Sept=medium.

If Sept=medium, then July=medium.

If August=medium, then Sept=medium.

If Sept=medium, then August=medium.

But, as backward prediction is not possible and invalid in the case of weather data, hence only three rules will be meaningful which are as shown below:

If June=high, then Sept=medium.

If July=medium, then Sept=medium.

If August=medium, then Sept=medium

(b) In this step confidence for the above frequent rules is calculated. Comparing the association with the predefined confidence threshold, denoted by λ as 0.5. Considering the first rule, the fuzzy count of the association "June_high ∩ Sept_medium” is found to be 3.6.

The confidence value of the rule “If June=high, then Sept=Medium” is given by

$$\sum_{1}^{10}(\text{June\_high} \cap \text{Sept\_medium}) / \sum_{1}^{10}(\text{June\_high}) = 0.9$$

In this way, the same process is done for other rules also, and

the result is as follows:

"If July=medium, then Sept=medium" has a confidence value of 0.96.

"If August=medium, then Sept=medium" has a confidence value of 0.78.

Step 10: After checking the predefined confidence threshold of 0.5 with the above confidence factor the following three association rules will be considered:

“If June=high, then Sept=medium.” with confidence 0.9.

"If July=medium, then Sept=medium" has a confidence of 0.96.

“If August=medium, then Sept=medium” has a confidence of 0.78.

4.3 FQARM:

The Fuzzy Quantitative Association Rule Mining Algorithm was proposed by A. Gyenesei [68] applied on quantitative dataset. It is an improvised form of the algorithm proposed by R. Srikant and et. al [76][77]. Here, the method of extraction of Fuzzy Association rules is classified into two parts:

Firstly, finding the frequent item sets by counting the fuzzy support and comparing them with user-specified minimum support.

Secondly, generating the rules from those frequent item sets

Simply it can be said, if X, Y, and XUY are frequent item sets i.e., it has more support count FS (Z, C) than the user-specified support value and high confidence value FC ((X, A), (Y, B)), then the rule X=>Y is said to be interesting. Here in this example, the support value and confidence value are assumed as 3.0 and 2.5 respectively.

To explain the algorithm, a sample dataset that consists of height, weight, and BMI_Status is undertaken as shown in Table 6. This technique deals with quantitative data and converts categorical indexes into binary form. The BMI (Body Mass Index) status can be quantified as 0 for unhealthy and 1 for healthy.

Table 6. Sample Dataset

Sl No	**Height (CM)**	**Weight (KG)**	**BMI_status**
1	173	95	Unhealthy
2	182	71	Healthy
3	172	68	Healthy
4	195	104	Unhealthy
5	158	45	Unhealthy
6	150	58	Unhealthy
7	175	65	Healthy
8	157	57	Healthy
9	150	40	Unhealthy
10	165	42	Unhealthy

Step1: Search (D): In this section of the procedure take the dataset and return the item set I= {i1, i2 i3………. In}. Here, in this example, I= {Height, Weight, BMI_Status} for the table 6.

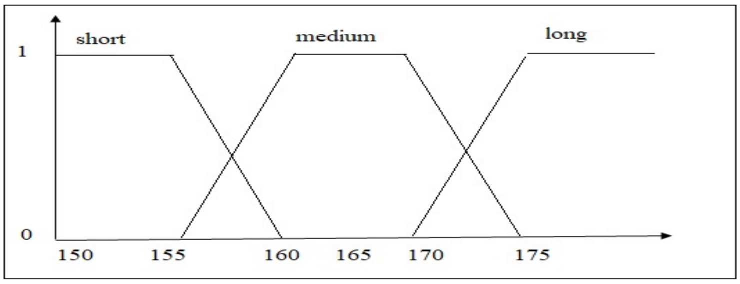

Figure 4. Membership Function used for height

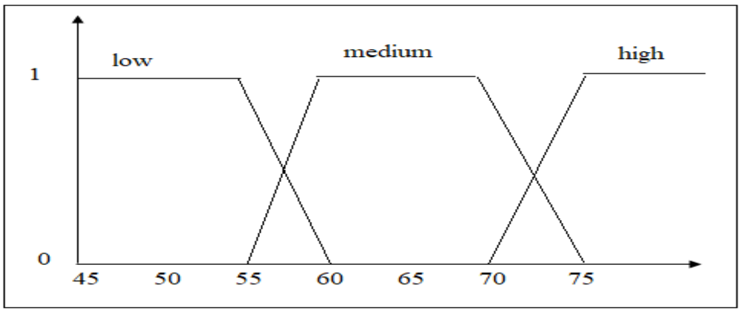

Figure 5. Membership Function used for weight

Step 2: Transform (D, T): In this step, dataset D is transformed to a fuzzy transaction dataset D^T by user-defined fuzzy sets. Parallelly, Candidate set C1 having 1-itemsets is found out from the updated database. The candidate set Ci includes sets of item and fuzzy set-like (item, fuzzy set) pairs. For example, C1= {{(Height, short)}, {(Height, medium)}, {(Height, long)}, {(Weight, low)} {(Weight, medium)} {(Weight, high)}is set of Candidate1-itemset. However, BMI_status is quantified as 1 or 0 for healthy or unhealthy respectively.

Table 7. Transaction dataset after fuzzification

Sl No	**Height (CM)**			**Weight (KG)**			**BMI_status**
	short	medium	long	low	medium	high	
1	0.0	0.4	0.6	0.0	0.0	1.0	0
2	0.0	0.0	1.0	0.0	0.8	0.2	1
3	0.0	0.6	0.4	0.0	1.0	0.0	1
4	0.0	0.0	1.0	0.0	0.0	1.0	0
5	0.4	0.6	0.0	1.0	0.0	0.0	0
6	1.0	0.0	0.0	0.4	0.6	0.0	0
7	0.0	0.0	1.0	0.0	1.0	0.0	1
8	0.0	1.0	0.0	0.6	0.4	0.0	1
9	1.0	0.0	0.0	1.0	0.0	0.0	0
10	0.0	1.0	0.0	1.0	0.0	0.0	0

Step 3: Checking (Ck, DT, minsup): Here, scanning the fuzzy dataset (DT), the fuzzy support (FS (X, A)) value of each candidate in Ck can be calculated. If the fuzzy support value of k-itemset is less than minsup, it is removed and frequent itemset Fk is extracted from Ck. Here, in this example, the minimum value is set as 3.0, and the minimum confidence is set as 2.5.

Step 4: Join (Fk-1): This step generates candidate set C2 from the frequent set F1 using the Join operation as follows:

Insert in to C2
Select {(X, A), (Y, B)}
From (X, A), (Y, B) in F1
Where X≠Y
Thus, after Join operation, C2 = {(Height, medium), (Weight, low)}, {(Height, medium), (Weight, medium)}, ….}, but C2 ≠ {(Height, medium), (Height, long)}, {(Weight, low), (Weight, medium)}, ….}. This Join step will be continued until the candidate set is generated.

Step 5: Prune (Ck): Here, if a subset of an item does not exist in Ck-1, then I in Ck will be pruned out

Step 6: Rules (F): Finally, association of frequent items is presented as rules.

Here, in this example, rules are found as follows:
If (Height, Medium) and (Weight, low) are frequent items, then the rule will be:
For the (BMI_Status, Healthy)
(Height, medium) => (Weight, low).
(Height, medium) => (Weight, medium).
(Height, long) => (Weight, medium).

4.4 CFARM

This algorithm operates on composite data, where multiple attributes are analysed to identify associations among those data items [69]. Composite data item can be defined as combination of several data [70][78]. Here, the algorithm is applied among nutrients in soil data to generate some rules. This soil health data is accessed from Krishi Vigyan Kendra, Dibrugarh, (Assam) under Assam Agriculture University, Jorhat. The quantity of macronutrients present in soil depends on the ingredients. Since, different ingredients can be found in a sample of soil with different quantities, hence to find possible associations with that composition, the Composite fuzzy association rule mining technique is very useful. Here, to work out the algorithm, a dataset of macronutrients of soil is prepared and then some useful rules are discovered using the CFARM algorithm.

Example: Here, as given below some key macronutrient composition of soil of farmer's land are shown. It is prepared as a raw dataset (D) which includes reading of different nutrients of the soil of different farmer's land. The unit micronutrient values are different and it is mentioned in Table 8.

Table .8. Sample data of different micronutrients of soil

Sl No	Land ID	pH (p^H)	OC (%)	Nitrogen (kg/ha)	Phosphorous (kg/ha)	Potassium (kg/ha)	Sulphur (Kg/ha)
1	A	5.6	0.7	448.9	16.1	147.3	32.5
2	B	5.1	1.0	556.9	18.9	145.8	35.7
3	C	5.8	0.6	557.8	21.1	123.9	34.7
4	D	5.4	0.8	501.9	20.1	137.9	31.1
5	E	5.8	0.8	467.9	22.8	143.1	23.6
6	F	5.9	0.9	436.9	23.1	125.8	22.1
7	G	4.9	0.7	345.9	18.9	136.6	32.5
8	H	5.0	0.9	459.9	21.1	143.1	35.7
9	I	5.2	0.9	509.8	20.1	129	34.7
10	J	5.7	0.7	476.7	22.8	142.9	31.9

The sample data consists of a set of micronutrient item I= {i_1, i_2, …., $i_{|I|}$} and properties P= {p_1, p_2, …., p_n}. The macronutrient data of some properties of soil (lands) is denoted as t_i[ij]= {v|v_1, v_2, …., vm}. If it is considered property value as "k^{th}" for j^{th} item in the i^{th} transaction, it can be shown as t_i[ij[vk]]. The notation <label, value> is used to represent each composite item of different lands of soil as given in Table 9.

Table 9. Example of raw dataset D

TID	Record
1	{<A, {5.6,0.7,448.9,16.1,147.3,32.5}>}
2	<B, {5.1,1.0,556.9,18.9,145.8,35.7}>
…..	………………………………………………………
10	{< J, {5.7,0.7,476.7,22.8,142.9,31.9}>}

As a part of the algorithm, initially, Property Dataset D^p is generated from raw dataset D. It consists of the property transaction (T^p) and property attributes (P). By aggregating the values for all Pj in t_1, the value for each property attribute can be computed as follows:

$$\text{Prop Value } (t_i^p[pj]) = \frac{\sum_{j=1}^{|ti|} ti[ij[vk]]}{|ti|}$$

Table 10. Example of properties dataset D^p

Micronutrients	Properties values
pH(I_1)	5.4
OC(I_2)	0.8
Nitrogen(I_3)	476.2
Phosphorus(I_4)	20.5
Potassium(I_5)	137.5
Sulphur(I_6)	31.4

In the second step, the property dataset is converted into the fuzzy dataset D´, which contains both fuzzy transaction (T´) and property attributes (P´). Subsequently, the fuzzy property attributes are partitioned into multiple linguistic labels L= {l_1, l_2, …, $l_{|l|}$}. Linguistic labels are classified into fuzzy membership degrees within the range between 0 to 1. Predefined fuzzy ranges are displayed in the properties table as given in Table 11.

Table 11. Fuzzy Intervals of property attributes table for the raw dataset.

Properties Attributes	Linguistic Labels		
	Low	Medium	High
pH(I_1)	$7.5 < V_k$	$8.5 \leq V_k \leq 4.5$	$V_k < 5.5$
OC(I_2)	$V_k<0.5$	$0.4 \leq V_k \leq 0.8$	$0.7< V_k$
Nitrogen(I_3)	$V_k<250$	$200 \leq V_k \leq 600$	$550< V_k$
Phosphorus(I_4)	$V_k<25.0$	$20.0 \leq V_k \leq 50.0$	$45.0< V_k$
Potassium(I_5)	$V_k<150$	$100 \leq V_k \leq 350$	$300< V_k$

Sulphur(I_6)	$V_k < 10$	$5 \leq V_k \leq 25$	$20 < V_k$

By employing the membership function μ ($t_i^p[pj]$, l_k), the numerical value of each property attribute $t_i^p[pj]$ is converted into membership degree. Consequently, the entire set of fuzzy properties is obtained through the product P and L. Table 12 displays the fuzzy dataset' derived from the property dataset.

Table 12 Fuzzy Dataset (T′)

Sl No	I1			I2			I3			I4			I5			I6		
	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
1	0	1	0	0	1	0	0	1	0	1	0	0	0.1	0.9	0	0	0	1
2	0	0.6	0.4	0	0	1	0	0.9	0.1	1	0	0	0.1	0.9	0	0	0	1
3	0	1	0	0	1	0	0	0.9	0.1	0.8	0.2	0	0.5	0.5	0	0	0	1
4	0	0.9	0.1	0	0	1	0	1	0	0.9	0.1	0	0.2	0.8	0	0	0	1
5	0	1	0	0	0	1	0	1	0	0.6	0.4	0	0.1	0.9	0	0	0.4	0.6
6	0	1	0	0	0	1	0	1	0	0.6	0.4	0	0.5	0.5	0	0	0.6	0.4
7	0	0.4	0.6	0	1	0	0	1	0	1	0	0	0.3	0.7	0	0	0	1
8	0	0.5	0.5	0	0	1	0	1	0	0.8	0.2	0	0.1	0.9	0	0	0	1
9	0	0.7	0.3	0	0	1	0	1	0	0.9	0.1	0	0.4	0.6	0	0	0	1
10	0	1	0	0	1	0	0	1	0	0.6	0.4	0	0.2	0.8	0	0	0	1

Then, an Apriori-based algorithm is used on the fuzzy dataset and some frequent items are generated based on user-specified threshold. The frequent items are listed in Table 13.

Table.13.Frequent Items

Itemset	Support
(I_3.M, I_4.L)	8.1
(I_3.M, I_6.H)	8.8
(I_4.L, I_6.H)	8.0

Again, this is based on the support count of possible item sets which can be found using the intersection operator between each of the two possible Item sets. Different combinations of frequent items can be found as shown in Table 14.

Table.14.Frequent Itemset

Items	Support
I1.M	8.1
I3. M	9.8
I4. L	8.2
I6.H	9.0

Thus, in the last step of the algorithm following associations in the form of some meaningful rules can be discovered:

If I_3=Medium then I_4=Low. (Support = 8.1)

If I_4= Low then I_3=Medium. (Support = 8.1)

If I_3=Medium, then I_6= High. (Support = 8.8)

If I_6= High, then I_3=Medium. (Support = 8.8)

If I_4=Low, then I_6=High. (Support = 8.0)

If I_6=High, then I_4=Low. (Support = 8.0)

Now, replacing the index by the corresponding property attribute, the following rules can be generated:

If Nitrogen=Medium then Phosphorous=Low. (Support = 8.1)

If Phosphorous=Low, then Nitrogen=Medium. (Support = 8.1)

If Nitrogen=Medium, then Sulphur=High. (Support = 8.8)

If Sulphur=High, then Nitrogen=Medium. (Support = 8.8)

If Phosphorous=Low, then Sulphur=High. (Support = 8.0)

If Sulphur=High, then Phosphorous=Low. (Support = 8.0)

Taking into consideration the pre-specified minimum confidence as 0.8 the measure of confidence is calculated for discovered rules. Thus, using the formula as given below, confidence for the first rule "If X_3=Medium then X_4=Low" can be calculated:

$\sum_1^{10}(\text{ If X3. Medium then X4. Low})/\sum_1^{10}\text{X3. Medium}$ =0.8

That is, If Nitrogen=Medium then Phosphorous=Low. (Support = 8.1) and confidence = 0.8.

Thus, the rules with their support count and confidence value are shown below:

If X_4= Low, then X_3=Medium. (Support = 8.1 & Confidence=0.9).

If X_3=Medium, then X_6=High. (Support = 8.8 & Confidence=0.8).

If X_6=High, then X_3=Medium. (Support = 8.8 & Confidence=0.9).

If X_4=Low, then X_6=High. (Support = 8.0 & Confidence=0.9).

If X_6=High, then X_4=Low. (Support = 8.0 & Confidence=0.8).

4.5 FWARM

The FWARM Algorithm [70] is a modified version of the Weighted ARM Algorithm. In WARM [79], a new concept was introduced assigning some weight to the item present in the transactional dataset. In classical ARM, each item present in the transaction was considered as equal significance 1. But it came to notice that some items may not be so important in generating interesting rules. In WARM, this issue was addressed by using weight to generate association rules on both Binary and Fuzzy Data. The FWARM Algorithm brought an improvised concept of assigning some Fuzzy weight against the item of collected dataset based on its significance or importance. The algorithm deals with some issues regarding the Downward Closure Property, that were not properly addressed by other quantitative or Boolean algorithms. The property is mainly based on the assumption that in the Apriori algorithm, all subsets are found to be large if the item is large. This way, DCP implies a large itemset by adding items to the itemset and increasing the size. But, in WARM algorithm, the DCP does not hold as each item is assigned by the weight and hence, an item may be large enough. In addition to this, the algorithm follows the breadth-first search traversal method which is based on tree data structures [80]. The execution process is as like as Apriori Algorithm and this algorithm also avoids pre-processing and post-processing steps to eliminate the additional steps. Here, we consider a dataset of different items with their Profit and Weight as a sample dataset to explain this approach.

Table. 15. Sample of weighted Item Dataset

ID	Items	Profit	Weight
A	Mobile	90	0.9
B	Laptop	70	0.7
C	Printer	50	0.5
D	Scanner	30	0.3

Consider a dataset D comprising transactions T= {t_1, t_2, …t_n} with itemset I= {i_1, i_2, …$i_{[l]}$}. A fuzzy dataset D' that consists of fuzzy transactions T'= $\{t_1', t_2', \ldots, t_n'\}$ is created. Each fuzzy set associates with item I. Items are identified with linguistic labels L {l_1, l_2, …., l_n} like {low, medium, high}. Each l in L associated with i is assigned with a weight w. Each attribute t_i' [i_j] is associated to some membership value ranging [0…1] with fuzzy sets. Thus "k^{th}" weighted fuzzy set for the "j^{th}" item in the "i^{th}" fuzzy transaction is given by t_i' [i_j[l_k[w]]] and each label l_k for item i_j would have associated with a weight. This pair ([i[l]], w) is called a weighted item where, [i[l]] €L. L is a label associated with i and w. Here, the following table shows the transaction having two fuzzy values of two linguistic term (low, medium).

Table. 16. Transaction Dataset

TID	Items
T1	A, B, C
T2	A, B
T3	B, C, D,
T4	B, D

Fuzzy Item Weight (FIW): FIW refers to a non-negative real value ranging from 0 to 1[0…1] as per the membership degree of each fuzzy set.

Fuzzy Itemset Transaction Weight (FITW): FITW refers to the aggregate weights of all fuzzy sets connected within items found in a single transaction.

Fuzzy Weighted Support: FWS can be computed by dividing the aggregated sum of FITW of all the item sets present in the transaction by the total number of transactions, denoted as:

$$FWS(X): \frac{sum\ of\ votes\ satisfying\ X}{Number\ of\ records\ in\ T}$$

Fuzzy Weighted Confidence: FWS is a ratio of some of the votes satisfying both XUY to the sum of votes satisfying X. It can be formulated as:

$$FWC(X \rightarrow Y): \frac{FWS(XUY)}{FWS(X)}$$

Step 1: Transform(T): In this step, the dataset that is considered for this example is converted into fuzzy transactional dataset.

Table. 17. Fuzzy Transaction Data

TID	X		Y	
	Low	medium	low	medium
T1	0.2	0.8	0.6	0.4
T2	0.4	0.6	0.8	0.2

T3	0.5	0.5	0.1	0.9
T4	0.7	0.3	0.5	0.5

Table. 18. Fuzzy Data with weight

Fuzzy Item I [l]	**Item Weight (IW)**
(X, low)	0.3
(X, medium)	0.5
(Y, low)	0.7
(Y, medium)	0.9

Step 2**:** Frequent Itemset generation(F): This subroutine generates frequent itemset from candidate item sets by comparing the FWS with pre-defined min_ws. To find FWS, FITW of an itemset (X, A) is calculated as:

Vote for t_i' satisfying X=$\prod_{k=1}^{|L|}(\forall[i[l[w]]] \in X)\ t_i'$ $[i_j[l_k[w]]]$

Here, from itemset < (X, low), (Y, medium)> we assume (X, low) as A and (Y, medium) as B. The FITW of itemset (A,

B) in transaction t_1 is calculated as:

FITW (A, B) = (0.3×0.2) × (0.9 × 0.4) = 0.0216.

FWS (A, B) $= \frac{\sum_{i=1}^{n}\prod_{k=1}^{|L|}(\forall[i[l[w]]]\in X)t_i' \text{ [ij[lk[w]]]}}{n} = \frac{0.2592}{4} = 0.0648$

FWC (X, Y) = $\sum_{i=1}^{n}\frac{\prod_{k=1}^{|Z|}(\forall[z[w]]\in Z)t_i' \text{ [Zk[w]]}}{\prod_{k=1}^{|X|}(\forall[i[w]]\in X)t_i' \text{ [Xk[w]]}}$, *where,* Z=A∪ B

FWC (A, B) = $\frac{WS(A\cup B)}{WS(A)} = \frac{0.0648}{0.135}$=0.48

Now, comparing the FWS with min_ws new frequent set can be generated as a candidate itemset.

Step 3: Rule Generation (R): in this step, potential rules (R) can be generated from frequent items based on the previous criteria

Step 4: Fuzzy Weighted Rule generation (R′): Finally, FAR can be generated by comparing the weighted confidence of candidate rules with a predefined confidence value.

5. Open Research Issues in FARM

FARM is considered one of the active research areas in data science because of its versatile application in different domains. With the increasing growth of data, extraction of useful and meaningful data has been identified as a common research challenge in recent decades. Although the progress of research is going on in various domains at different levels, it needs more attention to finding new and proper approaches. Based on the study of different algorithms of fuzzy association rule mining, some fundamental research issues can be considered for further research.

5.1 Membership function issue

Membership Function (MF) plays vital role in fuzzification of crisp data into fuzzy data. There is different type of membership functions used in fuzzification process. Some well-known membership functions are Triangular MF, Trapezoidal MF, Gaussian MF, Bell MF, singleton MF, etc. However, there is no appropriate assumption of selecting the suitable membership function. Here, based on this survey, it is found that the use of the membership function depends on the size and nature of the dataset. But more explanation is required to justify such an assumption.

5.2 Interval or range-defining issue

In the conversion of crisp data into fuzzy data there needs specific interval or range to categorize linguistic terms like low, medium, high, etc. In some cases, there are some standard ranges, based on which linguistic term can be classified. But, in some cases, no such ranges are found. In that case, ranges or intervals are classified based on assumption, which may lead variation in output.

5.3 Linguistic representation issue

Linguistic representation is a key feature of the FARM technique in the rule-generation process. It makes the output rules more user-friendly and effective. However, the problem arises when fuzzy data are classified. There is no convincing method that how many classifications can be done. Though, in some cases, it depends on the size of the dataset, there needs more justification.

5.4 Interestingness measures issue

Interestingness measures like Support, Confidence, and Lift are some common and well-established parameters used in the rule-generation process. These measures are predefined and based on that predefined value, frequent items are selected for the next iteration. This process is followed in most FARM approaches. However, inappropriate pre-assumed measures may lead to generating irrelevant or redundant rules. Assigning a proper value against such measures can be considered as a research issue to be addressed.

6. Summary and Discussion

The paper is based on a study of different algorithms related to FARM and its application in different fields. To perform the study, basic concept of fuzzy set theory and related applications in the last three years have been reviewed which is shown in Fig. 6. Out of different algorithms some key algorithms are selected for this study. The above-mentioned leading algorithms are based on distinct methodologies because of the variation in nature, methods, and dataset applied. There are various issues and challenges which are different from each other. Fuzzification is an important phase of these selected algorithms. The process of fuzzification mostly depends on membership function. The transformation of crisp value into fuzzy value differs according to the membership function. Different kinds of membership functions like triangular, trapezoidal, gaussian, etc can be used. As a result, both the fuzzification process and fuzzy values are different. Other differences can often arise in the classification of fuzzified data into linguistic terms. Linguistic terms may be varied depending on the dataset and ranges of classification. It is well established that the Apriori algorithm is the pioneer in mining associations among different attributes and presenting those associations in the form of rules. Evolution of fuzzy association rule mining algorithms begins with this Apriori-based algorithm and that is why almost every algorithm is based on the steps of that famous algorithm. Because of differences in nature, the application of such approaches is different. Despite these differences, there are also some similarities among those algorithms. Most of the algorithms are based on Apriori algorithm and support-confidence framework. Here, the summary is shown in Table 19 to identify the specialty of those algorithms for further work.

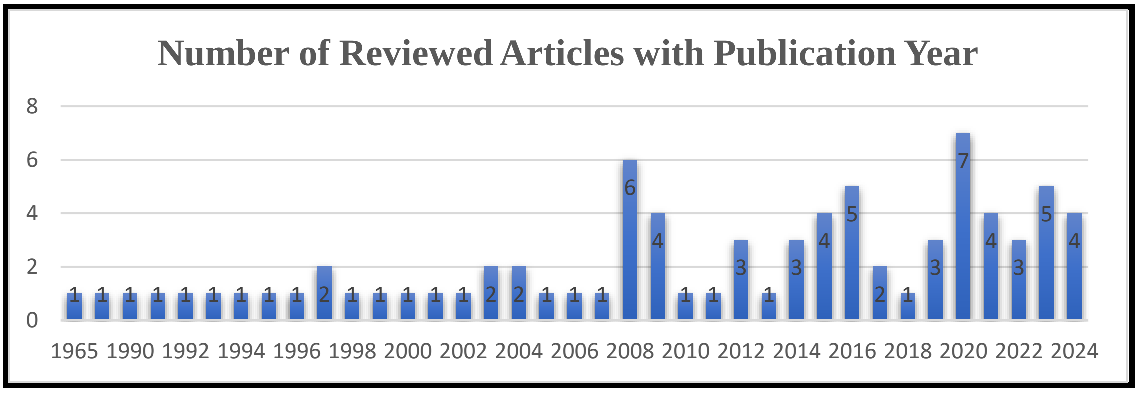

Fig. 6. Representation of the number of reviewed articles per year in the chart

Table.19. Summary of leading FARM Algorithms

Sl No	Name of the Algorithms	Authors	Nature	Data Type	Features /observations
1	Fuzzy Automatic Pattern Analysis and Classification System (F-APACS)	Keith C.C.Chan, Wai-Ho Au	FCM based	Categorical, Quantitative	Linguistic term is used instead of discretizing the domains of quantitative attributes. Adjusted difference is used in place of user supplied interesting measures. Discovery of both positive and negative association rules is possible.
2	Fuzzy Transaction Data Mining Algorithm (FTDA)	Tzung-Pei Hong, Chan-Sheng Kuo, Sheng-Chai Chi	Apriori TID-based	Quantitative, Categorical	A degraded membership function can be employed to solve conventional data-related problems. Smooth mining results with good time complexity can be attained when compared to traditional crisp set mining methods for quantitative data.
3	Fuzzy Quantitative Association Rule Mining (FQARM)	Attila Gyensei	Apriori-based	Quantitative, Binary	Sharp boundary problems can be solved using a fuzzy set. The fuzzy normalization process is used to tackle the problem raised in the fuzzy partition of quantitative data. The use of correlation measures helps to generate more accurate rules.
4	Composite Fuzzy Association Rule Mining (CFARM)	Maybin Muyeba, M. Sulaiman khan, Frans Coenen	Apriori-based	Composite data	A new concept of "composite items" is used which may link the related properties of a dataset and then find the associations among those attributes to make rules. The Use of the Certainty factor provides a good option to find a correlation among the rules.
5	Fuzzy Weighted Association Rule Mining (FWARM)	Maybin Muyeba, M. Sulaiman khan, Frans Coenen	Weighted-based	Quantitative	Based on importance or significance of data items, some value was assigned in weighted ARM and it leads an issue Downward Closure Property (DCP). The issue of invalidation of DCP is addressed within the support-confidence framework, applicable to both

					classical and fuzzy association rule.

7. CONCLUSION

This paper is based on a literature survey in the field of fuzzy association rule mining techniques. Here, different fuzzy-based algorithms for association rule mining proposed by different authors in the last three decades have been presented. The study includes the execution of those algorithms, findings, and key features. A summary of key feature those algorithms is mentioned. There are many possibilities and scope for research in this field to improve the performance of those algorithms in the future. Some basic issues are identified as the possible scope of research. With the tremendous growth of data taking place regularly, researchers, nowadays, need different methods of implementation to explore those data. As an output of this survey, some important issues related to predefined interesting measures, dataset representation, fuzzification process with proper membership function, and classical Apriori algorithm are found. The modification of this algorithm and its various issues can be considered for future work.

ACKNOWLEDGEMENT

The authors are thankful to the Department of Computer Science of Assam University, Silchar for providing the required infrastructure. The authors also acknowledged Prof. Prasanta Neog, Professor & HOD (Ag. Meteorology), PI (GKMS) BNCA, AAU, and Dr. Arup Kumar Sarma, Senior Scientist, Department of Entomology, AAU, Jorhat (Assam) for their kind help and valuable suggestion in collection and analysis of meteorological data, collected from BNCA Station under AAU, Jorhat (Assam) and Soil Health Card Data, collected from KVK, Dibrugarh, AAU, Jorhat (Assam).

REFERENCES

[1] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37-37.

[2] Mitra, S., & Acharya, T. (2005). Data mining: multimedia, soft computing, and bioinformatics. John Wiley & Sons.

[3] Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD international conference on Management of data (pp. 207-216).

[4] Delgado, M., Marín, N., Sánchez, D., & Vila, M. A. (2003). Fuzzy association rules: general model and applications. *IEEE transactions on Fuzzy Systems*, *11*(2), 214-225.

[5] Maeda, A., Ashida, H., Taniguchi, Y., & Takahashi, Y. (1995, March). Data mining system using fuzzy rule induction. In *Proceedings of 1995 IEEE International Conference on Fuzzy Systems.* (Vol. 5, pp. 45-46). IEEE.

[6] Yang, T., & Li, C. (2015, September). A study of fuzzy quantitative items based on weighted association rules mining. In *2nd International Conference on Intelligent Computing and Cognitive Informatics (ICICCI 2015)* (pp. 42-46). Atlantis Press.

[7] Sumathi, G., & Akilandeswari, J. (2020). Improved fuzzy weighted-iterative association rule-based ontology postprocessing in data mining for query recommendation applications. *Computational Intelligence*, *36*(2), 773-782.

[8] Allahverdi, N. (2019, October). Applications of fuzzy approach in medicine. problems and perspectives. In *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (pp. 1-8). IEEE.

[9] Han, J., Cai, Y., & Cercone, N. (1992, August). Knowledge discovery in databases: An attribute-oriented approach. In *VLDB* (Vol. 18, pp. 574-559).

[10] Dhanya, C. T., & Kumar, D. N. (2009). Data mining for evolving fuzzy association rules for predicting monsoon rainfall of India. *Journal of intelligent systems*, *18*(3), 193-210.

[11] Khan, M. S., Muyeba, M., Tjortjis, C., & Coenen, F. (2007). An effective fuzzy healthy association rule mining algorithm (FHARM). *databases*, *4*(5), 14.

[12] Das, P., Devi, L. P., & Gogoi, M. (2009). Nutrient composition of some regional recipes of Assam, India. *Studies on Ethno-Medicine*, *3*(2), 111-117.

[13] Sarma, R., & Sarma, P. K. D. (2020). Mining Composite Fuzzy Association Rules Among Nutrients in Food Recipe. In *Machine Learning, Image Processing, Network Security and Data Sciences: Second International Conference, MIND 2020, Silchar, India, July 30-31, 2020, Proceedings, Part II 2* (pp. 1-10). Springer Singapore.

[14] Chen, G., Yan, P., & Wei, Q. (2009). Discovering associations with uncertainty from large databases. *Recent Advances in Decision Making*, 45-66.

[15] Mangalampalli, A., & Pudi, V. (2012). FAR-miner: a fast and efficient algorithm for fuzzy association rule mining. *International Journal of Business Intelligence and Data Mining*, *7*(4), 288-317.

[16] Chan, K. C., & Au, W. H. (2001). Mining fuzzy association rules in a database containing relational and transactional data. *Data mining and computational intelligence*, 95-114.

[17] Pei, B., Zhao, S., Chen, H., Zhou, X., & Chen, D. (2013). FARP: Mining fuzzy association rules from a probabilistic quantitative database. *Information Sciences*, *237*, 242-260.

[18] Yan, P., Chen, G., Cornelis, C., De Cock, M., & Kerre, E. (2004, September). Mining positive and negative fuzzy association rules. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 270-276). Berlin, Heidelberg: Springer Berlin Heidelberg.

[19] Bhargava, R., Lade, S., & Panday, D. S. Mining Positive and Negative Fuzzy Association Rules with Item Cost. *Computer*, *90*, 0-9.

[20] Badhan B., JA Kabir M.M., “survey on association rule mining techniques. *published online*. (2019.May).

[21] Sundaravalli, N., & Geetha, A. (2016). A study & survey on rainfall prediction and production of crops using data mining techniques. *International Research Journal of Engineering and Technology (IRJET)*, *3*(12), 1269-1274.

[22] Buczak, A. L., Baugher, B., Guven, E., Ramac-Thomas, L. C., Elbert, Y., Babin, S. M., & Lewis, S. H. (2015). Fuzzy association rule mining and classification for the prediction of malaria in South Korea. *BMC medical informatics and decision making*, *15*, 1-17.

[23] Lee, C. K. H., Tse, Y. K., Ho, G. T., & Choy, K. L. (2015). Fuzzy association rule mining for fashion product development. *Industrial Management & Data Systems*, *115*(2), 383-399.

[24] Gürsel, G. (2016). Healthcare, uncertainty, and fuzzy logic. *Digital Medicine*, *2*(3), 101-112.

[25] Maheswari, D. U., & Gunasundari, R. (2017). User interesting navigation pattern discovery using fuzzy correlation based rule mining. *Int. J. Appl. Eng. Res*, *12*(22), 11818-11823.

[26] Kang, X., Porter, C. S., & Bohemia, E. (2020). Using the fuzzy weighted association rule mining approach to develop a customer satisfaction product form. *Journal of Intelligent & Fuzzy Systems*, *38*(4), 4343-4357.

[27] Rajest, S. S., Regin, R., & Shynu, T. (2023). DETECTION OF PROMINENT OBJECTS BY THE USE OF DEEP LEARNING. *CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES*, *4*(5), 62-82.

[28] Pan, T. (2021). An improved Apriori algorithm for association mining between physical fitness indices of college students. *International Journal of Emerging Technologies in Learning (iJET)*, *16*(9), 235-246.

[29] Sarno, R., Sinaga, F., & Sungkono, K. R. (2020). Anomaly detection in business processes using process mining and fuzzy association rule learning. *Journal of Big Data*, *7*(1), 5.

[30] Gupta, M. K., & Chandra, P. (2020). A comprehensive survey of data mining. *International Journal of Information Technology*, *12*(4), 1243-1257.

[31] Tasneem, T., Kabir, M. M. J., Xu, S., & Tasneem, T. (2023). A survey on data mining and machine learning techniques for diagnosing hepatitis disease. *International Journal of Biomedical Engineering and Technology*, *41*(4), 340-375.

[32] Ho, G. T., Ip, W. H., Wu, C. H., & Tse, Y. K. (2012). Using a fuzzy association rule mining approach to identify the financial data association. *Expert Systems with Applications*, *39*(10), 9054-9063.

[33] Vinodh, S., Prakash, N. H., & Selvan, K. E. (2011). Evaluation of agility in supply chains using fuzzy association rules mining. *International Journal of Production Research*, *49*(22), 6651-6661.

[34] Chatterjee, S., Chakrabarty, D., & Mukhopadhyay, A. (2022). Fuzzy association analysis for identifying climatic and socio-demographic factors impacting the spread of COVID-19. *Methods*, *203*, 511-522.

[35] Eligüzel, N. (2023). Analyzing society anti-vaccination attitudes towards COVID-19: combining latent dirichlet allocation and fuzzy association rule mining with a fuzzy cognitive map. *Fuzzy Optimization and Decision Making*, *22*(4), 669-696.

[36] Yue-Ju, X. U. E., Shu-Guang, L. I. U., Yue-Ming, H. U., & Jing-Feng, Y. A. N. G. (2010). Soil quality assessment using weighted fuzzy association rules. *Pedosphere*, *20*(3), 334-341.

[37] Pillai, J., Vyas, O. P., & Muyeba, M. K. (2014). A Fuzzy Algorithm for Mining High Utility Rare Itemsets-FHURI. *International Journal on Recent Trends in Engineering & Technology*, *10*(1), 1.

[38] Dogan, O., Kem, F. C., & Oztaysi, B. (2022). Fuzzy association rule mining approach to identify e-commerce product association considering sales amount. *Complex & Intelligent Systems*, *8*(2), 1551-1560.

[39] Dogan, O. (2023). A recommendation system in e-commerce with profit-support fuzzy association rule mining (p-farm). *Journal of Theoretical and Applied Electronic Commerce Research*, *18*(2), 831-847.

[40] Srivastava, D. K., Roychoudhury, B., & Samalia, H. V. (2019). Fuzzy association rule mining for economic development indicators. *International Journal of Intelligent Enterprise*, *6*(1), 3-18.

[41] Yang, T., & Li, C. (2015, September). A study of fuzzy quantitative items based on weighted association rules mining. In *2nd International Conference on Intelligent Computing and Cognitive Informatics (ICICCI 2015)* (pp. 42-46). Atlantis Press.

[42] Dong, X., Hao, F., Zhao, L., & Xu, T. (2020). An efficient method for pruning redundant negative and positive association rules. *Neurocomputing*, *393*, 245-258.

[43] Chen, C. Y., Liang, G. S., Su, Y., & Liao, M. S. (2014). A data mining algorithm for fuzzy transaction data. *Quality & Quantity*, *48*, 2963-2971.

[44] Matapurkar, P., & Shrivastava, S. (2020, June). Comparative analysis for mining fuzzified dataset using association rule mining approach. In *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)* (pp. 383-387). IEEE.

[45] Princy, S., & Dhenakaran, S. S. (2016). Comparison of triangular and trapezoidal fuzzy membership function. *J. Comput. Sci. Eng*, *2*(8), 46-51.

[46] Akgül, E., Delice, Y., Aydoğan, E. K., & Boran, F. E. (2022). An application of fuzzy linguistic summarization and fuzzy association rule mining to Kansei Engineering: a case study on cradle design. *Journal of Ambient Intelligence and Humanized Computing*, *13*(5), 2533-2563.

[47] Zhao, Z., Jian, Z., Gaba, G. S., Alroobaea, R., Masud, M., & Rubaiee, S. (2021). An improved association rule mining algorithm for large data. *Journal of Intelligent Systems*, *30*(1), 750-762.

[48] Lopez, F. J., Blanco, A., Garcia, F., Cano, C., & Marin, A. (2008). Fuzzy association rules for biological data analysis: a case study on yeast. *BMC bioinformatics*, *9*, 1-18.

[49] Kaya, M., Alhajj, R., Polat, 2., & Arslan, A. (2002). Efficient automated mining of fuzzy association rules. In *Database and Expert Systems Applications: 13th International Conference, DEXA 2002 Aix-en-Provence, France, September 2–6, 2002 Proceedings 13* (pp. 133-142). Springer Berlin Heidelberg.

[50] Chiu, H. P., Tang, Y. T., & Hsieh, K. L. (2012). Applying cluster-based fuzzy association rules mining framework into EC environment. *Applied Soft Computing*, *12*(8), 2114-2122.

[51] Yavari, A., Rajabzadeh, A., & Abdali-Mohammadi, F. (2021). Profile-based assessment of diseases affective factors using fuzzy association rule mining approach: A case study in heart diseases. *Journal of Biomedical Informatics*, *116*, 103695.

[52] Sarma, R., & Sarma, P. K. D. (2023, March). Fuzzy Association Rule Mining Techniques and Applications. In *International Conference on Advanced Computing, Machine Learning, Robotics and Internet Technologies* (pp. 73-87). Cham: Springer Nature Switzerland.

[53] Paiva, R. G., Cavalcante, C. A., & Do, P. (2024). Applying association rules in the maintenance and reliability of physical systems: A review. *Computers & Industrial Engineering*, *194*, 110332.

[54] Mirzakhanov, V. E. (2024). Fuzzy logic in association rule mining: limited effectiveness analysis. *Journal of Experimental & Theoretical Artificial Intelligence*, 1-15.

[55] Samsudin, N. A., & Matderis, M. (2024). Association Rules Mining Based on Fuzzy Soft Set.

[56] Tin, T. T., Hock, L. S., & Ikumapayi, O. M. (2024). Educational Big Data Mining: Comparison of Multiple Machine Learning Algorithms in Predictive Modelling of Student Academic Performance: Educational Big Data Mining. *International Journal of Advanced Computer Science & Applications*, *15*(6).

[57] Mangayarkkarasi, K., & Chidambaram, M. (2017). F-PNWAR: fuzzy-based positive and negative weighted association rule mining algorithm. *International Journal of Engineering and Technology*, *9*(6), 4250-4257.

[58] Mangayarkkarasi, K., & Chidambaram, M. (2018). E-FWARM: ENHANCED FUZZY-BASED WEIGHTED ASSOCIATION RULE MINING ALGORITHM. *Journal of Theoretical & Applied Information Technology*, *96*(2).

[59] Bao, F., Mao, L., Zhu, Y., Xiao, C., & Xu, C. (2021). An improved evaluation methodology for mining association rules. *Axioms*, *11*(1), 17.

[60] Zadeh, L. A "Fuzzy sets." *Information and control* 8.3 (1965): 338-353.

[61] Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning—I. *Information sciences*, *8*(3), 199-249.

[62] Pasi, G. (2008). Fuzzy sets in information retrieval: State of the art and research trends. *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, 517-535.

[63] Lu, J., Ruan, D., & Zhang, G. (2008). Fuzzy set techniques in e-Service applications. In *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models* (pp. 553-566). Berlin, Heidelberg: Springer Berlin Heidelberg.

[64] Pujari, A. K. (2001). *Data mining techniques*. Universities press.

[65] Agrawal, R. (1994). Fast Algorithms for Mining Association Rules. VLDB.

[66] K.CC. Chan and Wai-Ho Au. "Mining fuzzy association rules." *Proceedings of the sixth international conference on information and knowledge management*. 1997.

[67] Hong, T. P., Kuo, C. S., & Chi, S. C. (1999). Mining association rules from quantitative data. *Intelligent data analysis*, *3*(5), 363-376.

[68] Gyenesei, A. (2000, September). Mining weighted association rules for fuzzy quantitative items. In *European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 416-423). Berlin, Heidelberg: Springer Berlin Heidelberg.

[69] Khan, M. S., Muyeba, M., & Coenen, F. (2008). Mining Fuzzy Association Rules from Composite Items. In *Artificial Intelligence in Theory and Practice II: IFIP 20 th World Computer Congress, TC 12: IFIP AI 2008 Stream, September 7-10, 2008, Milano, Italy 2* (pp. 67-76). Springer US.

[70] Muyeba, M., Khan, M. S., & Coenen, F. (2009). Fuzzy weighted association rule mining with weighted support and confidence framework. In *New Frontiers in Applied Data Mining: PAKDD 2008 International Workshops, Osaka, Japan, May 20-23, 2008. Revised Selected Papers 12* (pp. 49-61). Springer Berlin Heidelberg.

[71] Chan, K. C., & Au, W. H. (1997, April). An effective algorithm for mining interesting quantitative association rules. In *Proceedings of the 1997 ACM symposium on Applied computing* (pp. 88-90).

[72] K.CC. Chan and A. KC Wong. "APACS: A system for the automatic analysis and classification of conceptual patterns." *Computational Intelligence* 6.3 (1990): 119-131.

[73] Chan, K. C. (1991). A statistical technique for extracting classificatory knowledge from databases. *Knowledge discovery in databases*.

[74] Hong, T. P., & Lee, Y. C. (2008). An overview of mining fuzzy association rules. *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, 397-410.

[75] Hong, T. P., Lin, K. Y., & Wang, S. L. (2003). Fuzzy data mining for interesting generalized association rules. *Fuzzy sets and systems*, *138*(2), 255-269.

[76] Sundaravalli, N., & Geetha, A. (2016). A study & survey on rainfall prediction and production of crops using data mining techniques. *International Research Journal of Engineering and Technology (IRJET)*, *3*(12), 1269-1274.

[77] Kuok, C. M., Fu, A., & Wong, M. H. (1998). Mining fuzzy association rules in databases. *ACM Sigmod Record*, *27*(1), 41-46.

[78] Wang, K., Liu, J. N., & Ma, W. M. (2006, December). Mining the Most Reliable Association Rules with Composite Items. In *Sixth IEEE International Conference on Data Mining-Workshops (ICDMW'06)* (pp. 749-754). IEEE.

[79] Khan, M. S., Muyeba, M., & Coenen, F. (2008). Weighted association rule mining from binary and fuzzy data. In *Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects: 8th Industrial Conference, ICDM 2008 Leipzig, Germany, July 16-18, 2008 Proceedings 8* (pp. 200-212). Springer Berlin Heidelberg.

[80] Coenen, F., Goulbourne, G., & Leng, P. (2004). Tree structures for mining association rules. *Data Mining and Knowledge Discovery*, *8*, 25-51.