Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Discovery of Fuzzy and Composite Fuzzy Association Rules in Meteorological Data

Rajkamal Sarma[1], Pankaj Kumar Deva Sarma[2]
[1]*Department of Computer Science, Assam University, Silchar*
[2] *Department of Computer Science, Assam University, Silchar*

ARTICLE INFO	ABSTRACT

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

Fuzzy Association Rule Mining (FARM) extends traditional ARM by evaluating and pruning rules based on interestingness measures to identify relevant patterns for various applications. The focus of this paper is to explore the application of FARM techniques demonstrating its algorithmic implementation in a meteorological dataset. Three major algorithms known as fuzzy Apriori, FTDA (Fuzzy Transaction Data-Mining Algorithm) and CFARM Composite Fuzzy Association Rule Mining) are experimented and analyzed. The experiment uses a real meteorological dataset spanning twenty years consisting some important attributes of weather such as rainfall, temperature, relative humidity, wind speed and bright sunshine hours of the North Bank Plain Zone (NBPZ) of the Brahmaputra River in Assam, India. The collected dataset is pre-processed into a transaction dataset and converted into a fuzzy dataset using membership functions. The three FARM algorithms are subsequently employed to uncover associations among various attributes within the fuzzy meteorological dataset. This study analyzes experimental results from three algorithms, focusing on factors like rule generation, computation time, and memory consumption. While Fuzzy Apriori provides comprehensive rule generation, it comes at the cost of higher computation time and memory usage. FTDA and CFARM, on the other hand, offer more efficient and significant rule generation, making them more suitable for large-scale, complex data analysis. The findings of this paper can contribute to the development of resilient and efficient data mining frameworks, enhancing the decision-making process for stakeholders in the meteorological domain. Thus, the paper introduces a new method for analyzing meteorological data using Fuzzy Association Rule Mining (FARM) techniques.

Keywords Data Mining, FARM, Apriori, FTDA, CFARM, Meteorological Dataset.

INTRODUCTION

Meteorological data analysis is considered as a meaningful practice in real life across various sectors and applications due to its pivotal role in informing decision-making, risk management, and resource allocation. Weather conditions profoundly impact various aspects of daily life, including agriculture, transportation, energy production, construction, public health, and disaster preparedness. Analysing meteorological data enables stakeholders to anticipate and respond to weather-related challenges, optimize resource utilization, and mitigate risks. For example, farmers rely on forecasts to plan planting schedules, irrigation strategies, and pest management to ensure an adequate food supply.

Transportation agencies use data to optimize routes and minimize disruptions, while energy sector decisions involve renewable energy generation and reducing fossil fuel reliance. Public health initiatives use meteorological data to enable early warning systems for heatwaves, air quality alerts, and disease outbreaks. Overall, meteorological data analysis contributes to societal well-being and resilience in the face of changing weather patterns and climate extremes.

Existing methods such as statistical methods, physical modelling, and machine learning approaches are employed in meteorological data analysis to investigate historical data, predict future weather patterns, simulate atmospheric interactions, and uncover complex patterns in large-scale datasets. Integrating these techniques provides a holistic

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

approach to weather analysis. The rapid expansion of data across various applications has posed challenges in uncovering valid and impactful association rules through data mining endeavours (Fayyad, et al., 1996). Data mining, a well-known data analysis technique, can be utilized to extract valuable insights from large and complex datasets generated by weather monitoring systems. Techniques such as association rule mining, clustering, classification, and regression reveal patterns, relationships, and trends, aiding in forecasting and decision-making in various sectors. Association rule mining, among different data mining techniques, can extract relationships among various weather parameters like rainfall, temperature, and humidity, aiding in weather forecasting, climate analysis, and decision-making by identifying rare or unexpected weather phenomena. The study (Ane, et.al., 2023) examines farming practices, the impact of technology on agriculture, and seasonal flower cultivation utilizing a flower dataset and data mining rules to analyze production values.

Fuzzy logic, a powerful computational tool used in mining association rules, handles uncertainty and imprecision in real-world datasets effectively. It allows for the gradual representation of linguistic variables and sets, enabling a nuanced analysis of data relationships. Fuzzy logic also incorporates domain knowledge and expert input, enhancing the interpretability and relevance of mined association rules. A fuzzy set encompasses a group of objects characterized by membership grades spanning from zero to one, which extend concepts such as inclusion, union, intersection, complement, relation, and convexity (Zadeh, 1965). The complete phenomena involve conventional linguistic operators that modify operand meanings, providing a means of approximate characterization for complex or ill-defined concepts that cannot be described quantitatively (Zadeh, 1975). Given the uncertainties and complexities inherent in meteorological data, fuzzy logic proves to be very effective in mining association rules in meteorological data analysis. Fuzzy Association Rule Mining (FARM) techniques offer a flexible approach to association rule mining compared to traditional methods. By leveraging fuzzy logic, FARM techniques capture gradual relationships between variables, enabling a nuanced analysis of complex datasets. Unlike traditional methods, which rely on crisp binary distinctions, FARM techniques accommodate uncertainty and imprecision present in real-world datasets. This flexibility allows FARM techniques to uncover subtle associations and dependencies that may be overlooked by traditional methods, making them highly effective in meteorological data analysis.

Rule mining is a key function in data mining practices, where rules are mined or generated from data to discover relationships among attributes of a transaction dataset. Association rule generation is significant in data mining and research activity, as it helps explain interesting relationships among various attributes (Agrawal, et al., 1993). In Market-Basket analysis, associations between items bought by customers are found from sales transactions. Association rules can be classified into different types like Boolean, generalized, and quantitative. However, these types have limitations in discovering nontrivial knowledge. Fuzzy set theory can be used to extend classical mining algorithms for databases containing values between 0 and 1, allowing for optimal representation of imprecise terms and relations (Delgado, et al., 2003). The simplicity of knowledge representation has led to the recognition of fuzzy-based techniques as an important component of data mining systems (Maeda, et al., 1995). The fuzzy concept in data mining is a methodology for extracting association rules from quantitative databases. It tackles the boundary problem encountered in attribute classification, which often arises from assuming a specific range of values, as seen in crisp sets where nearby values may be either overlooked or overly emphasized. Fuzzy sets alleviate this issue by assigning membership grades to multiple sets, particularly benefiting categorical data. Moreover, fuzzy sets effectively handle the partial membership of attributes in real-world scenarios by employing appropriate linguistic terms. Thus, the paper applies Fuzzy Association Rule Mining (FARM) techniques to a real-life meteorological dataset spanning twenty years, encompassing essential weather parameters such as rainfall, temperature, relative humidity, wind speed, and bright sunshine hours. Three rule mining algorithms, namely Fuzzy Apriori, FTDA, and CFARM, are implemented on the meteorological dataset to uncover associations among these key weather parameters. The resulting associations, represented as rules, are evaluated using performance measures to assess their significance in practical scenarios. By leveraging FARM techniques, this research aims to elucidate meaningful relationships and dependencies within the meteorological data, providing valuable insights for weather-related applications and decision-making processes.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

OBJECTIVES

The objective of this paper is to apply and analyze Fuzzy ARM algorithms on meteorological data to extract meaningful insights represented by association rules. These rules highlight the interdependencies among various weather parameters in the North Bank Plain Zone (NBPZ) of the Brahmaputra River in Assam. Such insights can be valuable for farmers in optimizing agricultural production and for researchers studying climate patterns and trends. Additionally, this paper intends to investigate the computational performance of three major FARM algorithms, comparing their efficiency and effectiveness in processing large meteorological datasets. Furthermore, the research seeks to discover both fuzzy and composite fuzzy rules to provide a more nuanced representation of weather patterns, capturing subtle variations that traditional rule-mining approaches might overlook.

LITERATURE RIVEW

Weather analysis is an active area for the researchers where data mining techniques are used to extract meaningful information. In this connection various weather prediction models were proposed by different researchers (Reddy, et al., 2017). In the paper, (Abhishekh, et al., 2018) an improved forecasting method using Intuitionistic Fuzzy Time Series (IFTS) with straightforward computational algorithms was introduced. The method fuzzifies historical time series data, resulting in distinct intuitionistic fuzzy relationship groups, offering a more accurate representation of forecast uncertainty and improved reliability. A comparison between the Apriori algorithm and Filter Associator in association rule analysis was done in weather dataset, focusing on frequent itemset generation and cycle performance (Baitharu, et al., 2015). In the study the rainfall patterns in Assam, Northeast India was analysed from 1981 to 2017 to investigate trends on a yearly, monthly, and seasonal basis (Gogoi and Rao, 2022). A study is performed on crop yield forecasting for rapeseed and mustard in the Brahmaputra valley of Assam, that used 25 years of yield and weather data (Kakati, et al., 2022). The findings showed that artificial neural networks (ANN) can accurately predict yield, with temperature and relative humidity being the most significant factors influencing yield across most districts. Association Rule Mining (ARM) was applied to identify latent patterns in climate data from 2013 to 2015, focusing on weather observations from Peta ling Jaya station in Selangor (Rashid, et al., 2017). Apriori is a well-established Association Rule Mining algorithm and have been improved time to time. An improvised form of the Apriori Tid algorithm was presented for association rule mining, focusing on rainfall data from North Eastern India, and analyzes performance and future directions (Sarma and Mishra, 2016). The use of multidimensional association rule mining across multiple transactions was explored, addressing challenges in efficient processing, and proposing an improved framework, a Modified Apriori algorithm, and its practicality in weather prediction (Nandagopal, et al., 2010). The research (Harun, et al., 2017) proposes a flood area prediction model using the Apriori algorithm on hydrological datasets from the Department of Irrigation and Drainage Malaysia, identifying villages with water levels across seven districts, and evaluating their significance using support, confidence, and lift values. An ensemble learning approach was introduced for improving rainfall prediction using multiple machine learning classifiers, focusing on Malaysian data and three algebraic combiners (Sani, et al., 2020). This paper (Chauhan and Thakur, 2014) reviews data mining techniques for weather prediction, comparing algorithms and findings. Decision tree and k-means clustering are found to be effective and have higher prediction accuracy compared to other methods. This study (Khan, et al., 2023) developed six multivariate models using past yield data and weather indices for three major soybean producing districts in Uttarakhand. This study (Seeboruth, et al., 2023) aims to develop a rainfall prediction model in New Delhi, India using the Mamdani fuzzy inference system. The model considers temperature, atmospheric pressure, humidity levels, dew point, and wind speed. This article (Paiva, et al., 2024) explores the role of descriptive data mining in improving maintenance and reliability in physical systems. It reviews association rules and their industrial applications, identifying research gaps. The article also highlights a surge in energy sector literature and proposes a research agenda for Industry 4.0, integrating climatic data with production processes and applying data mining in smart city infrastructure maintenance. This paper (Mirzakhanov, 2024) compares fuzzy and non-fuzzy association rule mining (ARM) effectiveness in associative classification. It reveals that fuzzy ARM can handle data inconsistencies, allowing classifiers to provide predictions and indicate certainty in output. The study also shows a correlation between classification accuracy and certainty, with greater certainty resulting in better classification accuracy.

Fuzzy Set theory

A fuzzy set builds upon the concept of a classical set, but it is distinguished by its membership function, which ranges continuously between 0 and 1, unlike the binary membership of classical sets constrained to 0 or 1. Formally, given a set X of elements x ∈X, any fuzzy subset A of X is defined as:

$$A= \{x, \mu A(x)|\ x \in X\} \quad (1)$$

Here, $\mu A(x):X|\rightarrow[0,1]$ represents the membership function in the fuzzy subset A, where the interval [0,1] is the range of real numbers between 0 and 1. The value μA(x) indicates the degree of membership of x in A, essentially expressing the degree to which x belongs to A.

Fuzzy set theory, pioneered by Lotfi Zadeh in the 1960s, revolutionized traditional binary logic. It introduced a more flexible and intricate approach to represent uncertainty and vagueness in data. Unlike classical set theory, where an element either belongs to a set or does not, fuzzy set theory allows for partial membership. For example: a student who scores 61% and another who scores 79% are both categorized as "Good," according to an evaluation system based on overall percentage of marks ranging from 60 to 80. However, this approach fails to capture the significant difference between their performances. Similarly, a student scoring 59% would be classified differently despite only a marginal difference from the "Good" category. This limitation, known as the sharp boundary problem, arises due to the binary nature of crisp sets. To address this issue, membership functions are very useful to represent values ranging from 0 to 1 using fuzzy logic, enabling a more detailed representation of data. Another scenario where fuzzy sets prove beneficial is in handling categorical data. For example, classifying items such as papayas and cucumbers as both fruits and vegetables using crisp sets would overlook their distinct characteristics. Fuzzy sets offer a solution by assigning membership grades to multiple sets, reflecting the overlapping nature of categories in real-world scenarios. Furthermore, binary association rules commonly used in data analysis assign Boolean values of 0 and 1 to indicate an attribute's participation. However, in many real-world situations, an attribute's membership to a set may be partial rather than absolute. By employing appropriate linguistic terms, fuzzy logic addresses this issue, enhancing the interpretability of the data and making it more accessible to human understanding. Thus, Fuzzy set theory enables the representation of gradual transitions between categories and captures the inherent uncertainty present in real-world data by using mathematical functions such as triangular, trapezoidal, or Gaussian curves. With applications across diverse fields such as control systems, pattern recognition, artificial intelligence, and decision-making, fuzzy set theory offers a powerful framework for modeling and reasoning with uncertain and imprecise information.

Review of Fuzzy ARM

Fuzzy ARM techniques are versatile tools used in various fields, including healthcare, CRM, financial services, manufacturing, environmental monitoring, transportation, and agriculture. They aid in medical data analysis, disease diagnosis, treatment recommendation, and patient prognosis, enhancing decision-making processes. CRM systems use FARM to analyze customer transaction data, enabling personalized marketing strategies and increased customer satisfaction. Financial services use FARM for fraud detection, risk assessment, and investment decision-making, leading to better risk management practices. Manufacturing optimizes production processes, while environmental monitoring predicts environmental phenomena like air quality and climate change. Transportation systems use FARM for traffic prediction, route optimization, and demand forecasting, leading to more efficient networks. Agriculture uses FARM for crop yield prediction, pest management, and irrigation optimization, improving productivity and sustainability. Since the formulation of the problem, commonly referred to as "the market-basket problem," researchers have devoted significant attention to uncovering meaningful associations among attributes in transactional data. [4]. Several algorithms are designed to discover the association rules applying different strategies. Algorithms such as Apriori, Apriori TID, etc., have been developed to enhance previous approaches in rule mining (Agrawal and Agrawal, 1994). However, the incorporation of fuzzy set theory has significantly transformed rule mining techniques and their associated algorithms. Such ARM algorithms typically involve the following key steps:

- Pre-processing of raw data involves transforming it into a sample dataset and presenting it as a transaction dataset.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

- The transaction dataset is then converted into a fuzzy dataset using membership functions, which assign degrees of membership to elements based on their similarity to predefined fuzzy sets.
- Fuzzified values are classified into linguistic terms to facilitate more intuitive and natural representation of the data.
- Frequent itemset generation is performed based on measures of interestingness, such as support and confidence, to identify patterns of association between items in the dataset.
- Finally, Rules are extracted from the frequent itemset, providing insights into the relationships and associations present in the data in a more natural and meaningful manner.

While the Apriori algorithm is acknowledged as a pioneer in ARM, recent decades have witnessed the development of several exciting algorithms in rule mining methods. Over the last three decades, algorithms such as F-APACS (Chan and Au, 1997), FTDA (Hong, et al., 1999), (Hong and Lee, 2008), FQARM (Gyenesei, 2000), CFARM (Khan, et al., 2008; Sarma and Sarma, 2020; Khan, et al., 2011) and FWARM (Muyeba, et al., 2009) have introduced new ideas to researchers, enhancing the efficacy of FARM techniques.

METHODOLOGY

To analysis the associations among different weather attributes, three algorithms are considered for this experimental work. The Apriori and FTDA algorithms are applied to fuzzy data, while the CFARM algorithm is utilized for composite fuzzy data analysis.

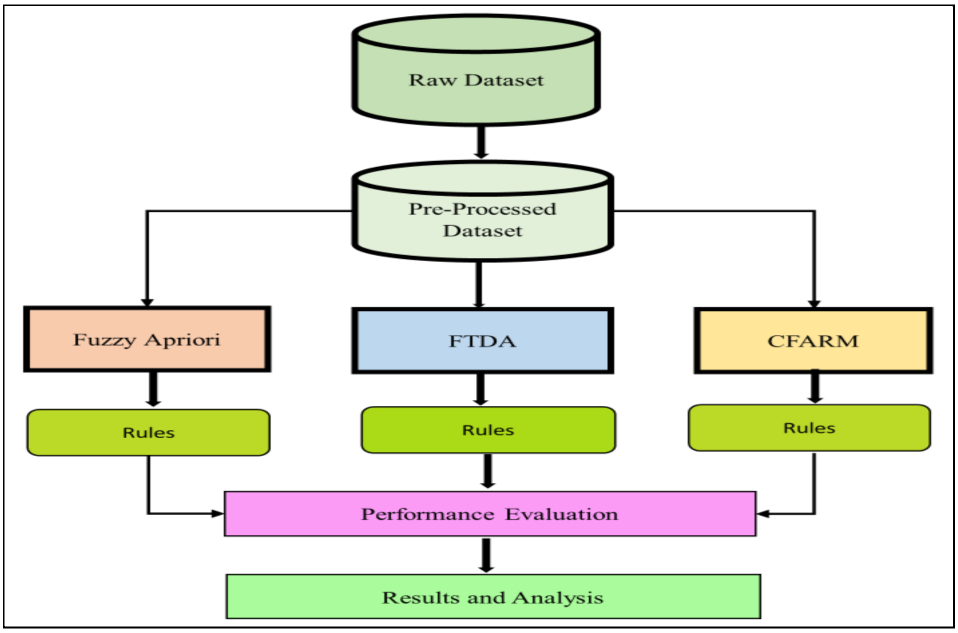

Figure 1: Process Flow Diagram

Data Collection and Description

The meteorological data utilized in this experiment originates from Biswanath College of Agriculture, Assam Agriculture University, Assam, India. The study location is situated in the NBPZ of the Brahmaputra River in Assam, encompassing six districts: Biswanath, Lakhimpur, Udalguri, Darrang, Sonitpur, and Dhemaji. The following map indicates the data location:

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

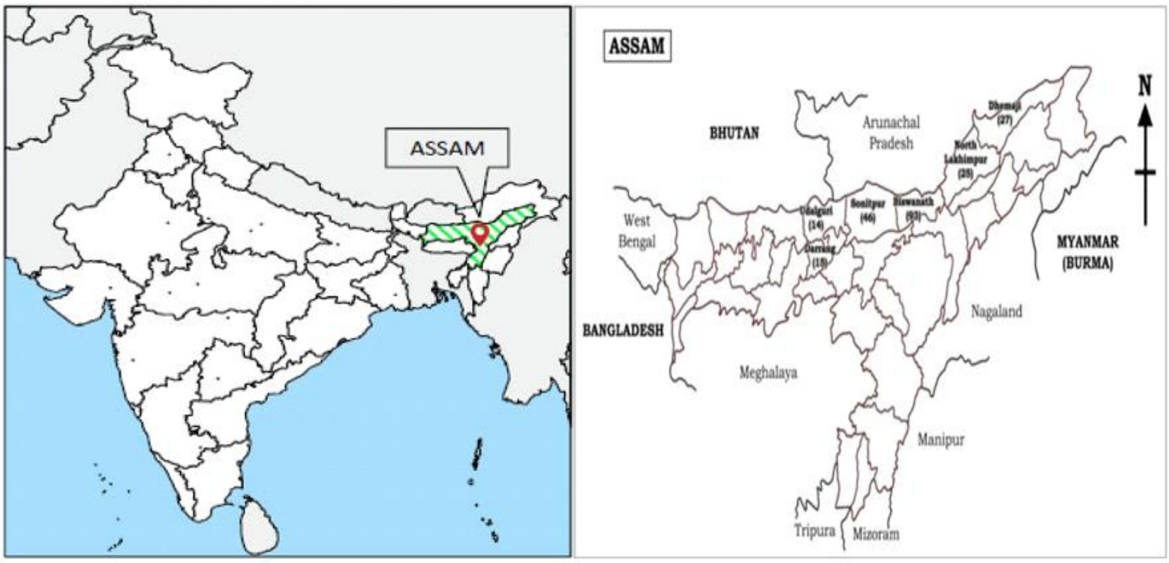

Figure 2 Map of NBPZ of Brahmaputra River of Assam, India, the data location for the study

The dataset spans a period of 20 years, from 2001 to 2020, and comprises five crucial meteorological attributes: Temperature, Relative Humidity, Rainfall, Wind Speed, and Bright Sun Shine Hours. It consists of 7305 daily transactions recorded from 1st January, 2001, to 31st December, 2020. However, in this experiment, meteorological data from the monsoon season spanning June to September has been preprocessed. The dataset is prepared for monthly basis by finding the average value of recorded attributes. A sample dataset prepared for the monsoon season during 2016 to 2020 is shown below:

Table 1: Sample weather dataset

ID	RF (mm)	TMP (°c)	RH (%)	WS (kmph)	BSSH (hrs)
Y16_06	10.3	29.2	80.4	2.9	5.2
Y16_07	12	28.3	85.2	3.1	2.8
Y16_08	8.7	30.6	77.7	2.8	6.1
Y16_09	8.7	28.8	82.9	2.5	4.7
Y17_06	15.7	28.1	82.8	3.2	3.4
Y17_07	10.3	29.1	81.7	3.3	4.6
Y17_08	10.6	29	83.7	2.4	3.9
Y17_09	9.6	28.8	83.9	2.1	4.3
Y18_06	10.2	28.8	81.4	3.1	4.3
Y18_07	13	29.1	84.3	2.4	4.2
Y18_08	11.2	29.2	82.7	2.6	4.7
Y18_09	6.5	28	74.5	2.2	4.7
Y19_06	6	28.3	84	2.6	4.9
Y19_07	12.2	27.6	87.2	2.5	3.2
Y19_08	3.4	29.9	81.6	2.3	6.3
Y19_09	10	27.2	86.6	2.2	4.2
Y20_06	14.1	27.3	85.2	3.1	3.3
Y20_07	9.2	28	87.8	2	2.3
Y20_08	5.7	29	83.5	2.1	5.6
Y20_09	8.1	27.9	86.8	1.8	3.4

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Data Preprocessing

Data preprocessing is a fundamental step in data mining that involves converting raw data into a suitable format for analysis. It addresses issues such as missing values, outliers, noise, and inconsistencies. The process includes data cleaning, data integration, data transformation, and data reduction. By performing data preprocessing, one can ensure data quality, reliability, and accuracy, leading to more meaningful insights and improved data mining tasks. Since the data considered for this work is real time data recorded on daily basis for twenty years long, hence it is very natural to occur some human error. In the collected dataset it is observed that there occur some empty values in some cases. In such cases, average value of that month is calculated. After doing needful correction the raw dataset is processed into transaction dataset.

Data Transformation

In this phase, the pre-processed transaction dataset undergoes a transformation into a fuzzy dataset by employing a trapezoidal membership function. This fuzzy dataset encompasses values expressed as fuzzy memberships, denoting their degree of participation in the set. The Membership function selection adheres to standard ranges established by the Indian Meteorological Department (IMD) for various attributes. While there exist several types of membership functions, including Triangular, Bell, Gaussian etc., the choice relies on the type of the dataset and its predefined ranges. In this particular scenario, a trapezoidal membership function is adopted due to the wide ranges of values within the transaction dataset.

The Trapezoidal curve is a function of a vector, x, and depends on four scalar parameters a, b, c, and d, as given below.

$$\mu(x) = \begin{cases} 0, & (x < a)\, or\, (x > d) \\ x - a/b - a, & a \le x \le b \\ 1, & b \le x \le c \\ d - x/d - c\,, & c \le x \le d \end{cases}$$

The parameters are defined by a lower limit a, an upper limit d, a lower support limit b, and an upper support limit c, where a < b < c < d

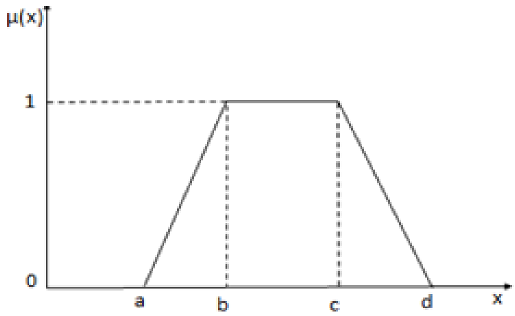

Figure 3: Trapezoidal Membership Function used for fuzzification

In the sample dataset, the temperature for June 2016 is recorded as 29.2°C. According to IMD standards, a medium temperature ranges from 15°C to 30°C, while temperatures above 30°C are classified as high. Therefore, in a crisp set, 29.2°C is considered a medium temperature. However, in a fuzzy set, which allows for a smooth boundary between ranges, the value can belong to both the medium and high categories. A fuzzy membership function transforms this crisp value into a fuzzy value by assigning membership degrees. For example, 29.2°C has a membership degree of 0.16 for medium temperature and 0.84 for high temperature. Similarly, other values are converted into fuzzy values with corresponding membership degrees.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 2: Fuzzy Dataset

YID	RF_ Low	RF_ Mid	RF_ High	TMP_ Low	TMP_ Mid	TMP_ High	RH_ Low	RH_ Mid	RH_ High	WS_ Low	WS_ Mid	WS_ High	BSSH_ Low	BSSH_ Mid	BSSH_ High
Y16_06	0	1	0	0	0.16	0.84	0	0	1	1	0	0	0.4	0.6	0
Y16_07	0	1	0	0	0.34	0.66	0	0	1	0.95	0.05	0	1	0	0
Y16_08	0	1	0	0	0	1	0	0	1	1	0	0	0	1	0
Y16_09	0	1	0	0	0.24	0.76	0	0	1	1	0	0	0.65	0.35	0
Y17_06	0	1	0	0	0.38	0.62	0	0	1	0.9	0.1	0	1	0	0
Y17_07	0	1	0	0	0.18	0.82	0	0	1	0.85	0.15	0	0.7	0.3	0
Y17_08	0	1	0	0	0.2	0.8	0	0	1	1	0	0	1	0	0
Y17_09	0	1	0	0	0.24	0.76	0	0	1	1	0	0	0.85	0.15	0
Y18_06	0	1	0	0	0.24	0.76	0	0	1	0.95	0.05	0	0.85	0.15	0
Y18_07	0	1	0	0	0.18	0.82	0	0	1	1	0	0	0.9	0.1	0
Y18_08	0	1	0	0	0.16	0.84	0	0	1	1	0	0	0.65	0.35	0
Y18_09	0.5	0.5	0	0	0.4	0.6	0	0	1	1	0	0	0.65	0.35	0
Y19_06	0.67	0.33	0	0	0.34	0.66	0	0	1	1	0	0	0.55	0.45	0
Y19_07	0	1	0	0	0.48	0.52	0	0	1	1	0	0	1	0	0
Y19_08	1	0	0	0	0.02	0.98	0	0	1	1	0	0	0	1	0
Y19_09	0	1	0	0	0.56	0.44	0	0	1	1	0	0	0.9	0.1	0
Y20_06	0	1	0	0	0.54	0.46	0	0	1	0.95	0.05	0	1	0	0
Y20_07	0	1	0	0	0.4	0.6	0	0	1	1	0	0	1	0	0
Y20_08	0.77	0.23	0	0	0.2	0.8	0	0	1	1	0	0	0.2	0.8	0
Y20_09	0	1	0	0	0.42	0.58	0	0	1	1	0	0	1	0	0

Based on Fuzzy membership degree the Fuzzy Dataset is transformed into Fuzzy Categorical Dataset using linguistic term Low, Medium and High as shown below in Table 3

Table 3: Fuzzy Categorical (Linguistic) Dataset

ID	RF	TMP	RH	WS	BSSH
Y16_06	Medium rainfall	High temperature	High humidity	Low wind	Medium sunshine
Y16_07	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine
Y16_08	Medium rainfall	High temperature	High humidity	Low wind	Medium sunshine
Y16_09	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine
Y17_06	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine
Y17_07	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine
Y17_08	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine
Y17_09	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine
Y18_06	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine
Y18_07	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine
Y18_08	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine
Y18_09	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine
Y19_06	Low rainfall	High temperature	High humidity	Low wind	Low sunshine
Y19_07	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Y19_08	Low rainfall	High temperature	High humidity	Low wind	Medium sunshine
Y19_09	Medium rainfall	Medium temperature	High humidity	Low wind	Low sunshine
Y20_06	Medium rainfall	Medium temperature	High humidity	Low wind	Low sunshine
Y20_07	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine
Y20_08	Low rainfall	High temperature	High humidity	Low wind	Medium sunshine
Y20_09	Medium rainfall	High temperature	High humidity	Low wind	Low sunshine

Data Mining

Data mining is a computational technique used to identify patterns, associations, and trends in large datasets. Its primary aim is to extract valuable insights from raw data, aiding organizations in enabling strategic decision-making and improving outcomes. The process involves stages such as data preparation, exploratory analysis, model development, assessment, and implementation.

Association Rule Mining

Association rule mining is a crucial aspect of data mining that uncovers patterns and relationships within large datasets. Association rule mining is a technique primarily used to discover patterns based on frequently occurring items within large transaction datasets. This method identifies correlations between items in transactional databases and represents them as rules in the form of X→Y, where X is known as the antecedent and Y as the consequent. The Apriori algorithm, a well-established method in this domain, is particularly effective at identifying frequent item sets and generating association rules.

Fuzzy Association Rule Mining

Fuzzy Association Rule Mining is one of the improvised forms of classical ARM approach. It is a technique that employs fuzzy set theory to deal with uncertainties in real-world data. It links items in transactional datasets with membership degrees, resulting in a more sophisticated representation of data relationships. The fundamental distinction between classical ARM and Fuzzy ARM is the method of association of data. While Classical ARM approach find association between crisp data, Fuzzy ARM deals with fuzzy data. Thus, it leads to the differences exist in classification and representation of classified data. It also addresses the sharp boundary issues present in classical ARM techniques by transforming crisp data into fuzzy data and converting quantitative datasets into fuzzy and binary categories (Sarma and Sarma, 2023). FARM is beneficial in fields such as data analysis, decision support systems, and pattern recognition to comprehend intricate data relationships. FARM techniques aim to enhance the accuracy and simplicity of associations among relevant attributes. The development of some FARM algorithms has led to more efficient rule generation, making it a promising approach for data mining applications. Here, in this paper, three existing algorithms are applied on same meteorological data to find the rules and to identify the methodical differences among those algorithms.

Performance Measures

Performance measures are crucial in ARM techniques for evaluating the quality and relevance of rules. They offer a quantitative basis for assessing how well rules capture significant patterns and correlations within a dataset. Key measures such as support, confidence, and lift determine the frequency and strength of associations between items. Support indicates how frequently an item set appears in a dataset confidence measures its reliability, and lift evaluates the independence of item sets. Some key performance measures are defined below:

Support: The support of an item set is the proportion of transactions in the dataset in which the item set appears. It indicates the frequency of occurrence of the item set in the dataset.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

$$Support(A) = \frac{Number of transactions containing A}{Total number of transactions} \quad (1)$$

Confidence: The confidence of a rule A -> B is the proportion of transactions containing A that also contain B. It measures how often items in B appear in transactions that contain A.

$$Confidence(A \to B) = \frac{Support(A \to B)}{Support(A)} \quad (2)$$

Lift: Lift is the ratio of the observed support to that expected if A and B were independent. It indicates the strength of the association between A and B. A lift value greater than 1 suggests a positive correlation.

$$Lift(A \to B) = \frac{Support(A \to B)}{Support(A) * Support(B)} \quad (3)$$

FARM Algorithms

Fuzzy Apriori Algorithm:

The Fuzzy Apriori algorithm is a modified version of the Apriori algorithm, specially designed for analyzing fuzzy datasets. It allows for partial membership in frequent item sets, providing a more nuanced analysis of itemset frequencies and association strengths. This algorithm is particularly useful in capturing gradual relationships and dependencies in fuzzy datasets, which may be overlooked by traditional binary approaches. The Fuzzy Apriori algorithm is applied on the meteorological Transaction dataset generated from Fuzzy categorial dataset shown in the table 4.

Table 4: Transaction Dataset

T1	**M_RF, H_TMP, H_RH, L_WS, M_BSSH**
T2	M_RF, H_TMP, H_RH, L_WS, L_BSSH
T3	M_RF, H_TMP, H_RH, L_WS, M_BSSH
T4	M_RF, H_TMP, H_RH, L_WS, L_BSSH
T5	M_RF, H_TMP, H_RH, L_WS, L_BSSH
T6	M_RF, H_TMP, H_RH, L_WS, L_BSSH
T7	M_RF, H_TMP, H_RH, L_WS, L_BSSH
T8	M_RF, H_TMP, H_RH, L_WS, L_BSSH
T9	M_RF, H_TMP, H_RH, L_WS, L_BSSH
T10	M_RF, H_TMP, H_RH, L_WS, L_BSSH
T11	M_RF, H_TMP, H_RH, L_WS, L_BSSH
T12	L_RF, H_TMP, H_RH, L_WS, L_BSSH
T13	M_RF, H_TMP, H_RH, L_WS, L_BSSH
T14	L_RF, H_TMP, H_RH, L_WS, L_BSSH
T15	M_RF, H_TMP, H_RH, L_WS, M_BSSH
T16	M_RF, M_TMP, H_RH, L_WS, L_BSSH
T17	M_RF, M_TMP, H_RH, L_WS, L_BSSH
T18	L_RF, H_TMP, H_RH, L_WS, L_BSSH
T19	M_RF, H_TMP, H_RH, L_WS, M_BSSH
T20	L_RF, H_TMP, H_RH, L_WS, L_BSSH

The algorithm is executed as the following steps:

Step 1: predefine of minimum support threshold - The support threshold is user defined measure calculated by the minimum number of times an item set must appear in a dataset to be considered frequent. This is typically determined by the user based on the size of the dataset and domain knowledge. Here, in this experiment, support count is set as 0.8.

Step 2: Generation a list of frequent 1-item sets - The complete dataset is scanned to detect items that qualify the minimum support threshold. These identified item sets are referred to as frequent 1-item sets.

Table 5: Frequent 1-Item set

Items	Support Count
M_RF(A)	0.8
H_TMP(B)	0.9
H_RH(C)	1.0
L_WS(D)	1.0
L_BSSH(E)	0.8

Step 3: Generation of candidate item sets - In this step, the algorithm generates a list of k+1 candidate item sets based on the frequent k-item sets identified in the previous step.

Table 6: Candidate Itemset-2

Candidate Itemset-2	Support Count
{A, B}	0.7
{A, C}	0.8
{A, D}	0.8
{A, E}	0.6
{B, C}	0.9
{B, D}	0.9
{B, E}	0.7
{C, D}	0.8
{C, E}	0.8
{D, E}	0.8

Step 4: Count the support of each candidate item set - The dataset is scanned again to count the number of times each candidate item set present in the dataset.

Table 7: Frequent Itemset-2

Candidate Itemset-2	Support Count
{A, C}	0.8
{A, D}	0.8
{B, C}	0.9
{B, D}	0.9
{C, D}	0.8
{C, E}	0.8
{D, E}	0.8

Step 5: Prune the candidate item sets - Item sets that do not meet the predefined support threshold are removed.

Table 8: Candidate Itemset-3

Candidate Itemset-3	Support Count
{A, C, D}	0.8
{A, C, B}	0.7
{A, C, E}	0.6
{A, D, B}	0.7
{A, D, E}	0.6
{B, C, D}	0.9

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

{B, C, E}	0.7
{B, D, E}	0.7

Table 9: Frequent Itemset-3

Candidate Itemset-3	Support Count
{A, C, D}	0.8
{B, C, D}	0.9

Step 6: The steps 3-5 are repeated until all possible frequent item sets are generated.

Step 7: Generation of association rules - After identifying the frequent item sets, the algorithm proceeds to find association from them in the form of rules. These association rules follow the form A→ B, where A and B represent item sets indicating if a transaction contains A, it is also likely to contain B.

Thus, possible association based on frequent itemset are:

{M_RF, H_RH, L_WS}	0.8
{H_TMP, H_RH, L_WS}	0.9

Step 8: Evaluation the association rules - Finally, the association rules are evaluated based on other metrics such as confidence and lift.

The Apriori property is a fundamental concept in the Apriori algorithm, stating that if an item set is frequent, its subsets must also be frequent. For instance, if an itemset {M_RF, H_RH, L_WS} frequently appears in a dataset, its subsets must also be frequent. The confidence and Lift of these association can be calculated by using equation (2) and (3) respectively.

$$\text{Confidence (M\_RF, H\_RH} \rightarrow \text{L\_WS)} = \frac{Support(M_{RF},H_{RH})\cup(L_{WS})}{Support(M_{RF},H_{RH})} = \frac{0.8}{0.8} = 1$$

$$Lift(M_{RF}, H_{RH} \rightarrow L_{WS}) = \frac{Support(M_{RF},H_{RH})\cup(L_{WS})}{Support(M_{RF},H_{RH}) * Support(L_{WS})} = \frac{0.8}{0.8*1.0} = 1$$

Finally, the rule "M_RF, H_RH→ L_WS" suggests that if "Rainfall is Medium" and "Relative Humidity is High", then "Wind Speed is Low" with both Confidence and Lift are absolutely 1,

In our experiment, employing Fuzzy Apriori algorithm on the fuzzy categorical dataset, several rules with measures of their interestingness have been achieved. The structure of patterns of generated rules representing those associations are different. Some associations occur between two items, while others involve more than two items. Thus, such associations provide more in-depth intendencies among the items

FTDA

The fuzzy Transaction Data-Mining Algorithm [20,21] is an improved algorithm specifically used in the FARM technique. This algorithm is utilized in quantitative datasets to identify significant associations in a fuzzy dataset. Unlike the Apriori algorithm, the fuzzy Transaction Data-Mining Algorithm does not directly execute the rule-mining process on the fuzzy dataset. Instead, it selects the most frequent fuzzy attributes from each of the original attributes. To determine the highest frequent attribute, it aggregates all the values of each column of the fuzzy dataset and compares each of the fuzzy categorical attributes to find the highest fuzzy categorical attribute. After selecting the most frequent fuzzy categorical attribute from each original attribute, mining process is executed to find the association based on different performance measures. The key steps of the complete algorithm are shown below:

Step 1: In the first step, the raw dataset is converted to a fuzzy dataset using a membership function. In our experiment conversion of sample dataset into fuzzy dataset using trapezoidal membership function is shown in the table 2.

Step 2: The membership degrees of each fuzzy categorical attribute (low, medium, high) are aggregated in this step to find the most frequent fuzzy category attribute among the original attributes. In the fuzzy dataset, there are three

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

linguistic representations like low, medium, and high against each attribute. In this step, fuzzy membership degree of each column is aggregated and we have found total frequency of each column with support count shown in following table 10.

Table 10: Transaction Dataset with Highest Frequent attributes

ID	RF_ Mid	TMP_ High	RH_ High	WS_ Low	BSSH_ Low
Y16_06	1.0	0.84	1.0	1.0	0.4
Y16_07	1.0	0.66	1.0	0.95	1.0
Y16_08	1.0	1.0	1.0	1.0	0.0
Y16_09	1.0	0.76	1.0	1.0	0.65
Y17_06	1.0	0.62	1.0	0.9	1.0
Y17_07	1.0	0.82	1.0	0.85	0.7
Y17_08	1.0	0.8	1.0	1.0	1.0
Y17_09	1.0	0.76	1.0	1.0	0.85
Y18_06	1.0	0.76	1.0	0.95	0.85
Y18_07	1.0	0.82	1.0	1.0	0.9
Y18_08	1.0	0.84	1.0	1.0	0.65
Y18_09	0.5	0.6	1.0	1.0	0.65
Y19_06	0.33	0.66	1.0	1.0	0.55
Y19_07	1.0	0.52	1.0	1.0	1.0
Y19_08	0	0.98	1.0	1.0	0.0
Y19_09	1.0	0.44	1.0	1.0	0.9
Y20_06	1.0	0.46	1.0	0.95	1.0
Y20_07	1.0	0.6	1.0	1.0	1.0
Y20_08	0.23	0.8	1.0	1.0	0.2
Y20_09	1.0	0.58	1.0	1.0	1.0
Total	17.06	14.32	20.0	19.6	14.3

Step 3: The most frequent fuzzy categorical attributes are identified, and the mining process is executed.

Table 11: Highest Frequent Attribute

RF_ Mid	**17.06**
TMP_ High	14.32
RH_ High	20.0
WS_ Low	19.6
BSSH_ Low	14.3

Step 4: In the mining process, the most frequent linguistic attribute from each attribute present in the fuzzy dataset is associated using the intersection operator. Then total membership of each association is calculated.

Table 12: Association between attributes

RF_ Mid	TMP_ High	RF_ Mid ∩ TMP_ High

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

1.0	0.84	0.84
1.0	0.66	0.66
1.0	1.0	1.0
1.0	0.76	0.76
1.0	0.62	0.62
1.0	0.82	0.82
1.0	0.8	0.8
1.0	0.76	0.76
1.0	0.76	0.76
1.0	0.82	0.82
1.0	0.84	0.84
0.5	0.6	0.5
0.33	0.66	0.33
1.0	0.52	0.52
0.0	0.98	0.0
1.0	0.44	0.44
1.0	0.46	0.46
1.0	0.6	0.6
0.2	0.8	0.2
1.0	0.58	0.58
TOTAL		12.31

Thus, possible associations with their total membership value are calculated.

Table 13: Possible association between attributes with aggregated support frequency

RF_ Mid ∩ TMP_ High	**12.31**
RF_ Mid ∩ RH_ High	17.06
RF_ Mid ∩ WS_ Low	16.66
RF_ Mid ∩ BSSH_ Low	17.93
TMP_ High ∩ RH_ High	14.32
TMP_ High ∩ WS_ Low	14.32
TMP_ High ∩ BSSH_ Low	10.77
RH_ High ∩ WS_ Low	19.6
RH_ High ∩ BSSH_ Low	14.3
WS_ Low ∩ BSSH_ Low	14.1

Step 5: The total frequencies of participation degree of the categorical attribute are aggregated, and these aggregated values are compared with a predefined performance measure support count. If it is greater than the given support count, then those attributes are selected for the next iteration, and others are pruned out from the mining process. Here, in this example, support count is assumed as 80%, that means value of minimum support count is 16. Hence, out of above all associations, the following association shown in the table, is qualified as frequent association. Now, based on associations previous table, new association can be generated as follows:

Table 14: Frequent Itemset-2

RF_ Mid ∩ RH_ High	**17.06**
RF_ Mid ∩ WS_ Low	16.66
RF_ Mid ∩ BSSH_ Low	17.93
RH_ High ∩ WS_ Low	19.6

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Step 6: As like previous steps, the possible association among attributes are generated as shown in the table 15 and based on predefined support count frequent sets are generated. Here, only one association can be found as shown in below and others are pruned out from the mining process.

Table 15: Candidate Set generation

RF_ Mid ∩ TMP_ High ∩ RH_ High	**12.34**
RF_ Mid ∩ TMP_ High ∩ WS_ Low	12.34
RF_ Mid ∩ TMP_ High ∩ BSSH_ Low	10.45
RF_ Mid ∩ RH_ High ∩ WS_ Low	16.66
RF_ Mid ∩ RH_ High ∩ BSSH_ Low	13.93
RF_ Mid ∩ WS_ Low ∩ BSSH_ Low	13.73
TMP_ High ∩ RH_ High ∩ WS_ Low	14.32
TMP_ High ∩ RH_ High ∩ BSSH_ Low	10.77
TMP_ High ∩ WS_ Low ∩ BSSH_ Low	10.77
RH_ High ∩ WS_ Low ∩ BSSH_ Low	14.1

From the above listed associations only following association is considered as frequent association based on predefined support count.

RF_ Mid ∩ RH_ High ∩ WS_ Low	16.66

Step 7: Steps 4-6 will be executed until all the frequent items are generated.

Step 8: Finally, the rules are generated, and other performance measures like confidence and lift are calculated. From above association different rules can be generated as stated below:

(RF_ Mid, RH_ High) → WS_ Low.

(RF_ Mid, WS_ Low.) → RH_ High.

(RH_ High, WS_ Low) → RF_ Mid.

The confidence and Lift of the rule (RH_ High, WS_ Low → RF_ Mid) can be calculated by using equation (2) and equation (3) respectively as given below:

$$\text{Confidence (RH\_ High, WS\_ Low} \rightarrow \text{RF\_ Mid)} = \frac{Support\{(RH_{High},WS_{Low})\cap(RF_{Mid})\}}{Support(RH_{High},WS_{Low})} = 0.85$$

$$Lift(RH_{High},WS_{Low} \rightarrow RF_{Mid}) = \frac{Support\{(RH_{High},WS_{Low})\cap(RF_{Mid})}{Support(RH_{High},WS_{Low})*Support(RF_{Mid})} = 0.049$$

Finally, rule (RH_ High, WS_ Low → RF_ Mid) suggest that if Relative Humidity is High and Wind Speed is low, then Rainfall is Medium in more 80% cases. The confidence of such rule is 85% and correlation is less than 1, which means association is negative association.

Thus, by employing the FTDA approach, fuzzy association rules can be achieved. The pattern of such rules presents an association among different attributes with the membership degree of those attributes. However, it is observed that compared to the Fuzzy Apriori algorithm, FTDA generates more significant rules.

CFARM Algorithm

The CFARM algorithm [26] is another well-established FARM algorithm used to find associations among different attributes present in a composite dataset. A composite dataset refers to a combination of several attributes, such as those found in food nutrients, soil nutrients, and medicinal compositions. The primary distinction of this algorithm from others is its process prior to converting the data into a fuzzy dataset. The average value of each attribute is calculated and stored as a property dataset. This property dataset typically holds the average value of each property attribute. Then, the property dataset is converted into a fuzzy dataset using a membership function. In order to create meaningful connections between fuzzy categorical property attributes, the defuzzification method is utilized. This method is based on the max-membership principle to identify the attribute that participates the most in the

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

categorical property. This approach helps to avoid generating unnecessary rules. The main steps of the CFARM algorithm are outlined below:

Step 1: Conversion of sample dataset into raw dataset. This can be done by grouping the YID, which represents year and month as A, B, C, D, E and displayed in table 1.

Table 16: Raw dataset

TID	RECORD
T1	[<a, (10.3, 29.2, …, 5.2)>, < b, (12, …., 2.8)>, <c, (8.7, …., 6.1)>, <d, (8.7, …., 4.7)>]
T2	[<a, (15.7, 28.1, …, 3.4)>, <b, (10.3, …, 4.6)>, <c, (10.6, …, 3.9)>, <d, (9.6, …, 4.3)>]
T3	[<a, (10.2, 28.8, …, 4.3)>, <b, (13, …., 4.2)>, <c, (11.2, …., 4.7)>, <d, (6.5, …, 4.7)>]
T4	[<a, (6, 28.3, …, 4.9)>, <b, (12.2, …, 3.2)>, <c, (3.4, …, 6.3)>, <d, (10, …., 4.2)>]
T5	[<a, (14.1, 27.3, …, 3.3)>, <b, (9.2, …., 2.3)>, <c, (5.7, …., 5.6), <d, (8.1, …., 3.4)>]

Step 2: Conversion of raw dataset into property dataset. This is calculated by finding average value of each attribute present in the raw dataset.

Table 17: Property Dataset

TID	YID	RF	TMP	RH	WS	BSSH
T1	Y16	9.93	29.23	81.55	2.83	4.70
T2	Y17	11.55	28.75	83.03	2.75	4.05
T3	Y18	10.23	28.78	80.73	2.58	4.48
T4	Y19	7.90	28.25	84.85	2.40	4.65
T5	Y20	9.28	28.05	85.83	2.25	3.65

Step 3: Conversion of property dataset into fuzzy dataset. The fuzzy dataset is transformed into linguistic or categorical form.

Table 18: Fuzzy dataset

YID	RF_Low	RF_Mid	RF_High	TMP_Low	TMP_Mid	TMP_High	RH_Low	RH_Mid	RH_High	WS_Low	WS_Mid	WS_High	BSSH_Low	BSSH_Mid	BSSH_High
Y16	0.0	1.0	0.0	0.0	0.15	0.85	0.0	0.0	1.0	1.0	0.0	0.0	0.65	0.35	0.0
Y17	0.0	1.0	0.0	0.0	0.25	0.75	0.0	0.0	1.0	1.0	0.0	0.0	0.98	0.02	0.0
Y18	0.0	1.0	0.0	0.0	0.24	0.76	0.0	0.0	1.0	1.0	0.0	0.0	0.76	0.24	0.0
Y19	0.03	0.97	0.0	0.0	0.35	0.65	0.0	0.0	1.0	1.0	0.0	0.0	0.67	0.33	0.0
Y20	0.0	1.0	0.0	0.0	0.39	0.61	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0

Step 4: Defuzzification of fuzzy dataset based on max-membership principle. This method is used to find significant associations between the fuzzy property attributes.

Table 19: Defuzzified Dataset

YID	RF_Mid (A)	TMP_High (B)	RH_High (C)	WS_Low (D)	BSSH_Low (E)
Y16	1	0.85	1.0	1.0	0.65
Y17	1	0.75	1.0	1.0	0.98
Y18	1	0.76	1.0	1.0	0.76
Y19	0.97	0.65	1.0	1.0	0.67
Y20	1	0.61	1.0	1.0	1.0

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Step 5: Implementation of Apriori Tree algorithm to find the rules. The fuzzy support of the selected fuzzy attributes is calculated by using the equation (1) and listed below:

Table 20: Average frequency of each attribute

Attributes	Support frequency
A	0.9
B	0.7
C	1.0
D	1.0
E	0.8

As per the Apriori algorithm support count is predefined as 80%, that means attribute "TMP_ High" is discarded for the next iteration. In the next iteration, frequent items are joined and new candidate set is generated. Again, the new candidate set are evaluated by given support count and item set having 80% of support count are advanced to the next iteration and infrequent sets are discarded. This process continues until all items are associated. The method of finding set of frequent attributes is like a tree structure as shown in the following figure:

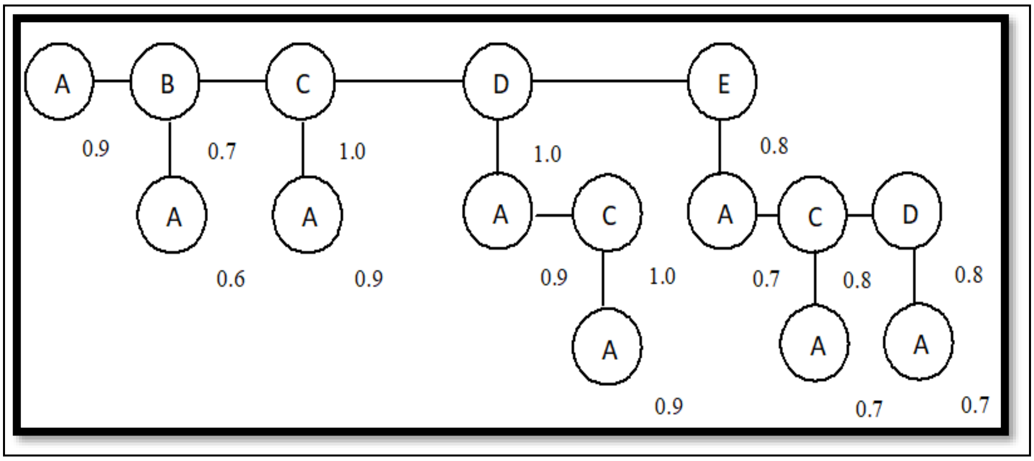

Figure 4: Structure of apriori Tree algorithm

Step 6: Generated rules are evaluated by different performance measures. Finally, frequent item set are presented in the rule form and those rules are evaluated by performance matrices like confidence and lift. Here, the fuzzy attribute set (A, C, D) is the only frequent set which means the set {RF_ Mid, Rh_ High, WS_ Low} is the frequent fuzzy attribute set based on predefined support count. Thus, the possible rules are

(RF_ Mid, RH_ High) → WS_ Low.

(RF_ Mid, WS_ Low.) → RH_ High.

(RH_ High, WS_ Low) → RF_ Mid.

The confidence and Lift of the rule {(RF_ Mid, WS_ Low.) → RH_ High} can be calculated by using equation (2) and equation (3) respectively as given below:

Confidence {(RF_ Mid, WS_ Low) → RH_ High} = $\frac{Support\{(RF_{Mid}, WS_{Low}) \cap (RH_{High})\}}{Support(RF_{Mid}, WS_{Low})}$ = 1.0

$$Lift\{(RF_{Mid}, WS_{Low}) \rightarrow RH_{High}\} = \frac{Support\{(RF_{Mid}, WS_{Low}) \cap (RH_{High})\}}{Support(RF_{Mid}, WS_{Low}) * Support(RH_{High})} = 1.0$$

Thus, CFARM algorithm is utilized in fuzzy ARM applications to extract associations from properties linked to composite attributes. The CFARM algorithm was modified to include a defuzzification method for generating significant rules, resulting in a smaller number of generated rules compared to the original CFARM algorithm.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

RESULTS AND DISCUSSION

The experimental work was conducted on a Windows 10 operating system, utilizing a system equipped with an 11th Gen Intel(R) Core (TM) i7-1135G7 @ 2.40GHz processor and 16 GB of RAM. The algorithms were implemented using the Python programming language. In this paper, three significant Fuzzy Association Rule Mining (FARM) algorithms are experimented on a meteorological dataset. The working methods of these algorithms are demonstrated using a sample dataset. This sample dataset is a subset of the original dataset, containing five key attributes: rainfall, temperature, relative humidity, wind speed, and bright sunshine hours. The dataset spans the monsoon seasons from 2016 to 2020, as defined by the Indian Meteorological Department (IMD), which includes the months of June, July, August, and September. For the sample dataset, monthly average data for these months is calculated. The fuzzy Apriori algorithm employed in this experiment is an extension of the traditional Apriori algorithm. The execution process of the Fuzzy Apriori algorithm is similar to that of the traditional Apriori algorithm, with the key difference being that the Fuzzy Apriori algorithm is applied to a fuzzy dataset instead of a standard transaction dataset. The Fuzzy Apriori algorithm's output shows two types of associations among three frequent item sets. One association is among Moderate Rainfall, High Relative Humidity, and Low Wind Speed, and another association is among High temperature, High Relative Humidity, and Low Wind Speed. On the other hand, with respect to FTDA and CFARM, only one type of association is found. Although all three algorithms generate almost similar kinds of associations, the execution process differs for each one. Fuzzy Apriori and FTDA algorithms are suitable for fuzzy datasets, while CFARM is suitable for composite fuzzy datasets. The frequent itemset generation method also varies in each algorithm. While Fuzzy Apriori is employed to find the number of occurrences of items in the transaction dataset, the other two algorithms emphasize on finding fuzzy values rather than occurrences. In FTDA, the highest frequent fuzzy attribute is selected based on its maximum aggregated values from each attribute. In the instance of the CFARM algorithm, the average of those four months against each attribute is calculated to transform the transaction dataset into the property dataset, and then fuzzification is done. Thus, it requires less computation time than the other two algorithms in the fuzzification of transaction datasets. Some key observations are found out based on the method and experimented results as given below.

Effect on number of generated rules:

The rule generation process is inversely proportional to the predefined values of support and confidence across all three algorithms. As the support and confidence thresholds increase, the quantity of derived rules decreases. Compared to the Fuzzy Apriori algorithm, FTDA and CFARM produce fewer rules. However, the rules generated by FTDA and CFARM tend to be more significant. The following figure shows the rule generation trends of these algorithms

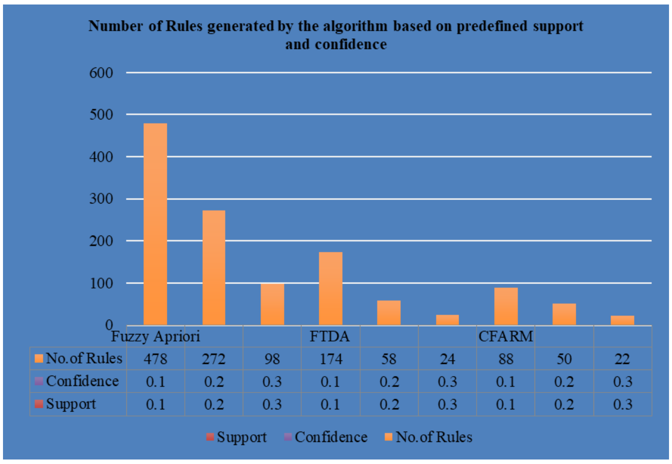

Figure 4: Graphical representation of number of generated rules based on predefined Support and Confidence.

Effect on computation time

The efficiency of these algorithms in the rule generation process can be analyzed by examining their computation times. Using the support-confidence framework, the computation times for each algorithm are illustrated in the

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

following figure. Compared to FTDA and CFARM, the Fuzzy Apriori algorithm requires considerably more time in the rule generation process.

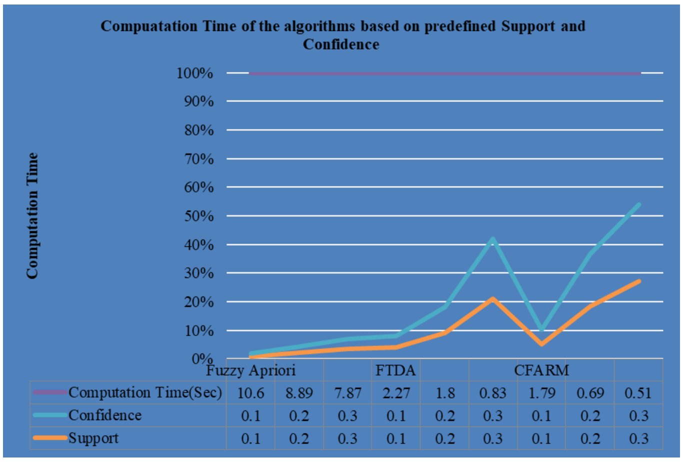

Figure 5: Graphical representation of trends of required computation time w.r.t. predefined support and confidence

Effect on memory consumption

Memory consumption is a key aspect of the rule mining process. In the experiment, the memory usage of the three algorithms was computed based on different values of support and confidence. The following graph visualizes the memory consumption trends of the different algorithms. Once again, the Fuzzy Apriori algorithm requires more memory, whereas CFARM uses the least memory in the rule mining process

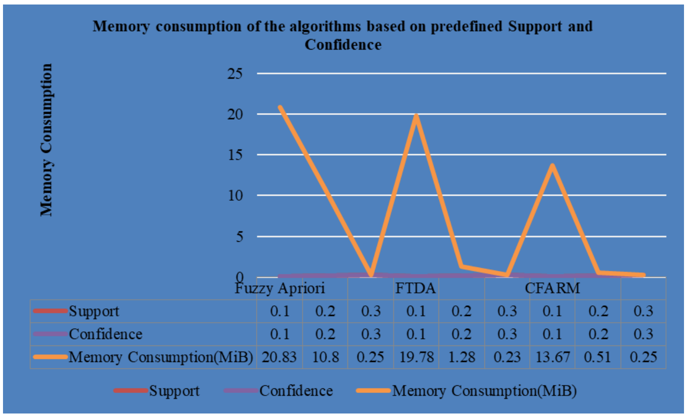

Figure 6: Graphical representation of memory consumption of algorithms based on support and confidence

The outcomes of the experiment provided key insights into the capabilities of these algorithms in the rule generation process, including the quantity of generated rules, computation time, and memory consumption. Fuzzy Apriori, an extension of the classical Apriori algorithm, is easy to execute in contrast to the other algorithms. However, it requires longer duration for frequent itemset generation as it extracts more rules. The FTDA algorithm selects the highest frequent attribute after fuzzification, which discards irrelevant attributes directly and reduces the steps needed for frequent itemset generation. Consequently, it requires less computation time and memory. The CFARM algorithm enhances the FARM technique by mining rules in composite data. In this algorithm, the average value is calculated to convert the transaction dataset into a property attribute dataset before fuzzification. Hence, an aggregated average value is considered for fuzzification. This approach enables the extraction of more significant rules with a smaller amount of execution time and memory utilization. Overall, each algorithm has unique strengths in finding associations among attributes. However, FTDA and CFARM demonstrate greater efficacy and accuracy in the rule mining process

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

CONCLUSION

In this study, step by step process of three major Fuzzy Association Rule Mining (FARM) algorithms used in a meteorological dataset are explained. Initially, the dataset is pre-processed into a workable form using various techniques. Analyzing meteorological data is inherently challenging due to the numerous methods available. While various statistical methods have been traditionally used in meteorological data analysis, data mining remains a reliable approach for deriving valuable insights across various fields. Association rule mining, a key part of data mining, enhances the process in multiple ways. Introducing fuzzy concepts into ARM extends the rule mining process, providing greater accuracy and generating more in-depth information.

The algorithms considered for this experiment—Fuzzy Apriori, FTDA, and CFARM—are well-established and have been used in various domains. In this paper, we apply them to a meteorological dataset, potentially aiding decision-making for relevant stakeholders in this field. Our contribution includes modifying the CFARM algorithm with defuzzification to enhance the significant rule generation process. Thus, this research can offer a novel method in forecasting and pattern recognition by associating different parameters for meteorological data analysis

REFRENCES

[1] 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).
[2] Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB (Vol. 1215, pp. 487-499).
[3] Abhishekh, Gautam, S. S., & Singh, S. R. (2018). A score function-based method of forecasting using intuitionistic fuzzy time series. New Mathematics and Natural Computation, 14(01), 91-111.
[4] A. Rashid, R. A., Nohuddin, P. N., & Zainol, Z. (2017). Association rule mining using time series data for Malaysia climate variability prediction. In Advances in Visual Informatics: 5th International Visual Informatics Conference, IVIC 2017, Bangi, Malaysia, November 28–30, 2017, Proceedings 5 (pp. 120-130). Springer International Publishing.
[5] Ane, T., Nepa, T., & Khan, M. R. (2023). Smart and intelligent production strategy for the flower market using data mining knowledge-based decision. Bangladesh Journal of Multidisciplinary Scientific Research, 7(1), 35-43
[6] Baitharu, T. R., & Pani, S. K. (2015). A comparative study on associaton rule mining algorithms using weather dataset. International Journal of Advanced Technology in Engineering and Science, 3, 291-295.
[7] Chauhan, D., & Thakur, J. (2014). Data mining techniques for weather prediction: A review. International Journal on Recent and Innovation Trends in Computing and Communication, 2(8), 2184-2189.
[8] Chan, K. C., & Au, W. H. (1997, January). Mining fuzzy association rules. In Proceedings of the sixth international conference on information and knowledge management (pp. 209-215).
[9] 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.
[10] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37-37.
[11] Gogoi, K., & Rao, K. N. (2022). Analysis of Rainfall Trends over Assam, North East India. Curr World Environ, 17(2), 435-446.
[12] 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.
[13] Harun, N. A., Makhtar, M., Abd Aziz, A., Zakaria, Z. A., & Syed, F. (2017). The application of apriori algorithm in predicting flood areas. management, 17, 18.
[14] Hong, T. P., Kuo, C. S., & Chi, S. C. (1999). Mining association rules from quantitative data. Intelligent data analysis, 3(5), 363-376.
[15] 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.
[16] Khan, Y., Kumar, V., Setiya, P., & Satpathi, A. (2023). Comparison of phenological weather indices based statistical, machine learning and hybrid models for soybean yield forecasting in Uttarakhand. Journal of Agrometeorology, 25(3), 425-431.
[17] Khan, M. S., Muyeba, M., & Coenen, F. (2008). A Framework for Mining Fuzzy Association Rules from Composite Items. ALSIP (PAKDD).

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

[18] Khan, M. S., Muyeba, M., Coenen, F., Reid, D., & Tawfik, H. (2011). Finding associations in composite data sets: the CFARM algorithm. International Journal of Data Warehousing and Mining (IJDWM), 7(3), 1-29.
[19] Kakati, N., Deka, R. L., Das, P., Goswami, J., Khanikar, P. G., & Saikia, H. (2022). Forecasting yield of rapeseed and mustard using multiple linear regression and ANN techniques in the Brahmaputra valley of Assam, North East India. Theoretical and Applied Climatology, 150(3), 1201-1215.
[20] 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.
[21] 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.
[22] Mirzakhanov, V. E. (2024). Fuzzy logic in association rule mining: limited effectiveness analysis. Journal of Experimental & Theoretical Artificial Intelligence, 1-15.
[23] Nandagopal, S., Karthik, S., & Arunachalam, V. P. (2010). Mining of meteorological data using modified apriori algorithm. European Journal of Scientific Research, 47(2), 295-308.
[24] Paiva, R. G., Cavalcante, C. A. V., & Do, P. (2024). Applying association rules in the maintenance and reliability of physical systems: A review. Computers & Industrial Engineering, 110332.
[25] Reddy, P. C., & Babu, A. S. (2017, February). Survey on weather prediction using big data analytics. In 2017 Second international conference on electrical, computer and communication technologies (ICECCT) (pp. 1-6). IEEE.
[26] Sarma, H. K. D., & Mishra, S. (2016, December). Mining time series data with Apriori tid algorithm. In 2016 International Conference on Information Technology (ICIT) (pp. 160-164). IEEE.
[27] Sani, N. S., Abd Rahman, A. H., Adam, A., Shlash, I., & Aliff, M. (2020). Ensemble learning for rainfall prediction. International Journal of Advanced Computer Science and Applications, 11(11).
[28] 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.
[29] 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.
[30] Seeboruth, K., Wen, L. Z., Hameed, V. A., Ling, T. Y., Rajadorai, K. P., & Rana, M. E. (2023, December). Fuzzy Logic Approach to Predicting Rainfall Patterns. In 2023 IEEE 21st Student Conference on Research and Development (SCOReD) (pp. 432-436). IEEE.
[31] Zadeh, L. A. (1965). Fuzzy sets. Information and control, 8(3), 338-353.
[32] Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning—I. Information sciences, 8(3), 199-249.