

A Hybrid Integrated Model for Big Data Applications Based on Association Rules and Fuzzy Logic: A Review

Hind R. Ibraheem¹, Murtadha M. Hamad¹

University of Anbar, Ramadi, Iraq

Corresponding author: Hind R. Ibraheem

DOI: <https://doi.org/10.52866/ijcsm.2023.02.02.015>

Received December 2022 ; Accepted March 2023 ; Available online April 2023

ABSTRACT: There is a real increase in the generation of data from different sources. Data mining is a useful method to elicit valuable information. Association rule mining can assist in finding patterns and trends in big data. Also, fuzzy logic plays a main role as an assistance technique in handling big data issues. This review paper present recent literature on the hybridization of association rule mining or other data mining techniques such as classification and clustering and fuzzy logic techniques in big data. Whereas a hybrid model of association rule and fuzzy logic is suggested to get valuable knowledge for big data applications at good accuracy and less time, with the aid of distributed frameworks for big data handling (Hadoop, Spark, and MapReduce). Different techniques and algorithms were used in these works and evaluated according to accuracy, sensitivity, recall, and run time with various results such as sensitivity = 80%, specificity = 86% ,and F-measure = 2.5, or achieving high accuracy and shorter runtime compared to other methods and 98.5 accuraccy of fitness function in pruning redundant rules . At the end of the paper, we present the most prominent and widely used techniques that assist in providing useful and valuable knowledge in different domains from huge, unstructured, and even heterogeneous data. The paper will be beneficial to researches who are interested in the field of mining big data.

Keywords: Big Data, Association Rule Mining, Fuzzy Logic, Distributed Framework.

1. INTRODUCTION

Recently, many DMs and machine learning (ML) algorithms were suggested for extracting knowledge from big data. the prominent vs. features, like veracity, value, velocity, variety, and volume, are used to describe such data, which is referred to as big data in literature. Value is particularly concerned with the knowledge that might be mined from data that is valuable. Ad-hoc programming paradigms like MapReduce have typically been used to build such algorithms on particular distributed computing frameworks like Apache Spark and Apache Hadoop [1]. The Latest DM approaches heavily rely on association rules. An association rule is demonstrated when two related products are purchased together. Depending on how the programmer sets up the system for DM, fuzzy logic might be obvious or take the form of an estimate when forecasting the next set of data [2]. The use of fuzzy in DM will enable programmers to create specific values without being aware of their exact value. The standard logic, which states that a program can only specify a speed as either slow or fast, is fundamentally different from this [3]. The structure of this essay is as follows. An introduction to the ideas of big data characteristics, association rule mining algorithms, platforms, and fuzzy logic methodologies is provided in Section 2. The association rules mining using fuzzy logic and big data have been illustrated in Section 3 and in some related works. The topics covered in these studies are discussed and analyzed in Section 4. Finally, in Section 5, we reach a decision.

2. BACKGROUND

Below different and important terms will show in a brief manner. The upcoming sections will demonstrate big data, association rule mining and fuzzy logic concepts.

2.1 Big Data Characteristics

Big data is the term that is utilized in order to describe the rapid expansion of heterogeneous data flows as a result of the widespread use of new technologies. Using websites, social networks, and mobile devices expanding daily, information is currently more readily available than ever. The term "extensive information" refers to the enormous

volume of semi-structured, structured, and unstructured data that could be usefully mined for the information [4]. Big data is typically characterized using the five Vs [5], [6]: -

- Volume: This translates to a huge volume of data that fluctuates between terabytes and zettabytes each second. Those enormous datasets could be handled by distributed databases.
- Velocity: The maximal rate of throughput is a reflection of the volume of data generated or transmitted during flow processing.
- Variety: which exemplifies the enormous variety of information that is available.
- Veracity: Recognizing data efficiency and accuracy.
- Value: This demonstrates that valuable information is discovered from data.

2.2 Big Data Platforms

A big data platform is a kind of IT solution which combines the functions and features of many big data tools and applications into a single solution. Those platforms consist of Spark, Hadoop, and MapReduce.

- Originally created in the early 2000s, Hadoop Big Data can be defined as one of the big data platforms which enables businesses to process, store, and analyze massive data collections. Hadoop Big Data has emerged as a potent tool to process and manage data as a result of the growth of big data. [7].

- Massive scalability over hundreds to thousands of the servers in Hadoop cluster has been made possible through MapReduce programming model. Apache Hadoop's core is MapReduce, which serves as the processing component. [8].

- Unified analytics engine for big data processing called Spark is open-source. With implicit data parallelism and fault tolerance, Spark offers a programming interface for clusters. Spark was first created at the University of California, and its resilient distributed dataset (RDD) serves as its architectural cornerstone [9].

2.3 Association Rule Mining

According to co-occurrence features, association rules mining (ARM) could discover potential relations between variables from large datasets with no prior knowledge. It's a common strategy in the retail sales sector when a business is looking to uncover products that are regularly bought in tandem. Recently, association rule approaches have primarily fallen into one of three categories [10]:

2.3.1 Apriori Algorithm

The Apriori algorithm is a method for discovering frequent item sets by employing association rule learning on the dataset and identifying data patterns. This algorithm, which is frequently utilized in market basket analysis, needs a bigger dataset [11]- [12].

Advantages of Apriori:

- Depth first search reduces memory requirements and no need to scan of $(k+1)$ itemsets for $K \geq 1$.

Disadvantage of Apriori:

- TID-sets could be quite long, making manipulation expensive.

2.3.2 FP-growth Algorithm

The Apriori method has been improved with the use of this algorithm. Without any necessity for the candidate generation, a frequent pattern has been created. The frequent pattern tree (FP tree) or FP growth algorithm uses a tree for representing the database.

Advantage of FP-Growth:

- Only 2 pass over dataset. No candidate generation.; Compresses dataset, and much faster than Apriori.

Disadvantage of FP-Growth:

- FP-Tree might not fit in Memory; Building an FP-Tree is expensive.

2.3.3 Eclat Algorithm

Eclat algorithm utilizes a vertical database layout, where each one of the items is stored along with its cover (also known as Tidlist) rather than directly listing all transactions. It then utilizes an intersection-based method to calculate the support of an item set. It is more time-efficient than apriori pattern generation and ideal for small datasets [13]- [14].

2.4 Fuzzy logic

When it is complicated to characterize states as "False" or "True," the fuzzy logic system is used. Partially false and partially true values could be included in the fuzzy logic. It attempts to resemble human decision-making that might take under consideration all of the values between True and False. The architecture diagram that follows explains the four main components of fuzzy logic system [15]:

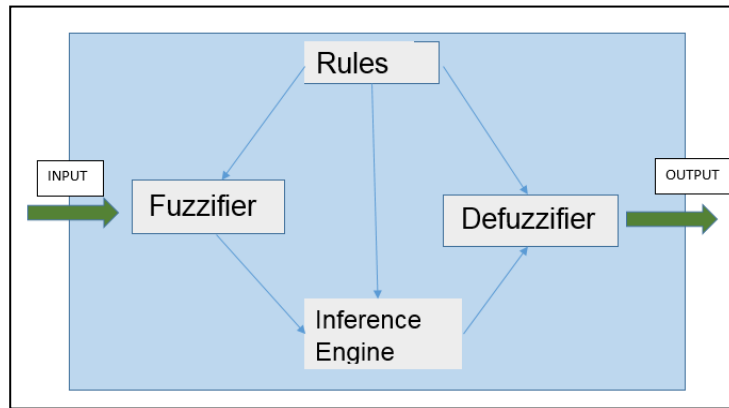


Figure.1 fuzzy logic diagram

3. LITERATURE SURVEY

In this section some of literature survey of several researchers will be display which help in this study that depend on association rules mining, fuzzy logic techniques, data mining techniques and big data.

- **Vinod, et. al. (2015)** [16], In order to forecast the risks associated with diagnosing diseases like breast cancer, the study makes an effort to suggest a fuzzy association algorithm. The work's findings showed that the prediction is dependable compared to traditional approaches. Age, menopause with value being 'premeno', node-caps being 'no', degree of malignancy being are factors that influence the diagnosis of cancer.
- **Segatori, et. al. (2017)** [17], This paper suggests a useful MapReduce-based distributed fuzzy associative classification technique. The method makes effective use of a new distributed discretizer depending on the fuzzy entropy to generate fuzzy partitions of characteristics. After that, using a distributed fuzzy variant regarding well-known FP-Growth algorithm, a set of potential fuzzy association rules is produced, and applying three different forms of pruning on this set. 11,000,000 instances were used in experiments with a focus on computation time, accuracy and scalability.
- **Wang, et. al. (2017)** [18] When processing big data, fuzzy set approaches are crucial since they not only model algorithmic solutions' uncertainties but also the granules of information they contain. In a variety of contributions, fuzzy set approaches are used in big data applications. A few of them adapt the current fuzzy set-based algorithms to big data environment, whereas others work to create new methods that are appropriate for tackling certain big data problems.
- **Almukahel, et. al. (2018)** [19], A novel version of MapReduce with four layers is created by combining MapReduce and fuzzy logic in a hybrid method. The first layer is used to collect data. Preprocessing data is the second layer, in which semi-structured data is cleaned up and given relations. Fuzzy controllers and classification are applied in the third layer to provide rules. Data reduction and classification are lastly applied in the fourth and final layer in order to produce a significant and predictive result.
- **El-Mandouh, et. al. (2018)** [20], with the use of well-known conventional K-means clustering algorithm, this work suggests a hybrid approach to cluster big data. The three-phased suggested methodology divides big data into smaller datasets in the first phase using the map-reduce algorithm. The typical clustering K-means algorithm is used to every one of small datasets that were separated in the second phase. The final stage is in charge of generating the general clusters output for the entire data set. The final phase involved the implementation and comparison of two functions, Mode and Fuzzy Gaussian, to choose the most appropriate one.

- **Osuwa, et. al. (2018)** [21], the study will discuss the state-of-the-art pattern-based DM data before demonstrating how fuzzy logic can be used to analyze, structure, and forecast future data. Without requiring much time, data could be recovered from trillions of pieces of information. The results will demonstrate that fuzzy logic could effectively detect the data on what the user wants, and there are additional benefits to applying it in DM.
- **Golnar Assadat and Afzali,Shahriar Mohammadi. (2019)** [22], in addition to its advantages in revealing ambiguous linkages between data, ARM will violate privacy. To prevent sensitive association rules from being found, association rule hiding could be helpful. Fuzzy logic-based new big data association rule hiding method is provided. In the suggested approach, characteristics like scalability and parallelism are embedded.
- **Ducange, et. al. (2020)** [23], this paper focuses specifically on classification applications and aims to provide an overview regarding distributed algorithms for learning fuzzy models from the Big Data. As was to be predicted, only distributed versions of algorithms for creating fuzzy models could effectively handle dataset sizes greater than 0.5 GB. The emphasis on the works where the experiments made use of actual big datasets.
- **Madbouly, et. al. (2021)** [24], in order to discover ARs that attain minconf and minsup values, the proposed method uses a hybridization of Type 2 fuzzy logic system T2FLS with ARs to know which item sets are frequently encountered. After 30 generations, it reaches a fitness value of 98.52%, however the optimization of the machining rules requires the most time, almost 2.988 seconds. The rule has a lot of accuracy potential because it improves non-redundant FARs with the use of GA and is more accurate when redundant rules are removed

Table 1: Comparison of studies that contained feature selection, association rules, fuzzy logic and big data technologies

No.	Citation	Techniques and Algorithms Used	Aim	Result	Dataset
1.	Lekha,A.,C.V.Srikri shna and V. Vinod 2015[16]	Applying fuzzy logic on association rule mining as well as Clustering techniques	The proposed method aim to assessment and diagnoses of cancer risks.	The results shows the effect of several attributes to the infection of breast cancer disease	Obtained from University Medical Centre, Oncology Institute, Ljubljana, Yugoslavia.
2.	Armando Segatori et al. 2017 [17]	Basic concepts on the fuzzy associative classifications and MapReduce distributed programming model.	The work aim To enhance consistent computational complexity and memory occupation.	Achieving high accuracy and shorter runtime compared to other methods	the algorithm tested on 6 well-known big datasets, which have been extracted from UCI5 and LIBSVM6 repositories
3.	Hai Wang, Zeshui Xu and Witold Pedrycz. 2017 [18]	Fuzzy set techniques	This work is about handling big data in an effective manner which results in intelligent decision making.	They infer based on some principles possible trends of using fuzzy sets in big data processing	Various datasets from UCI.
4.	Ikhlas Almukahel, Wael Alzyadat and Mohamad Alfayomi 2018[19]	Hybrid method utilizing the fuzzy logic and MapReduce is utilized for producing a new MapReduce version	The model aim to enhance Pre-processing and analysis of big data to obtain the valuable information	Results have shown the efficiency of the method via Specificity = 86%, Sensitivity = 80%, and F-measure= 2.5	R (Readr, Dplyr, PreProcess, HadoopStreaming, TidyR, FuzzyR and HiveR).In addition to Weka.

<p>5. Amira M. El-Mandouh et al 2018 [20]</p>	<p>K-means clustering algorithm, map-reduce algorithm to split big data to small data-sets. Mode and Fuzzy Gaussian were implemented.</p>	<p>The proposed hybrid model aim to clustering big data with the using of K mean clustering algorithm and Map Reduce framework.</p>	<p>The outcomes demonstrated the model's effectiveness while clustering big amounts of data with the K-means algorithm. Additionally, the tests demonstrate that Fuzzy Gaussian function outperforms the conventional Mode function in terms of accuracy.</p>	<p>Experimental study utilized 4 benchmark big datasets, which include; Covtype, Poker, Covtype-2, and Poker-2.</p>
<p>6. Abdulhafis A.Osuw, Abdulazeez Akande Osuwa and Abdulazeez A. 2018 [21]</p>	<p>Fuzzy logic, pattern techniques and clustering.</p>	<p>The DM system used in this study employs pattern recognition and fuzzy logic. Review the fuzzy logic and the existing DM systems.</p>	<p>The findings will demonstrate that fuzzy logic might precisely identify the data on what the user wants. There are further benefits to employing fuzzy logic in DM.</p>	<p>Database of chemical compounds (concentration of substances in mg/l)</p>
<p>7. Golnar Assadat and Afzali,Shahriar Mohammadi.. 2019 [22]</p>	<p>A new big data association rule hiding technique (data anonymisation and parallelization) is presented, uses fuzzy membership degree approach.</p>	<p>The system aim to avoid the risk of sensitive knowledge leakage. And decreasing unwanted side effect on non- sensitive rules</p>	<p>The proposed method minimized unwanted side effect of hiding sensitive rules on non-sensitive rules</p>	<p>Two dataset are used: Brijs dataset collected from 1999 to 2000 and Clue Web dataset were collected in 2009</p>
<p>8. Pietro Ducange,Michela Fazzolari and Francesco Marcelloni 2020 [23]</p>	<p>Generation of the fuzzy classification models, Fuzzy Rule-Based Classifiers and Fuzzy Decision Trees like the Fuzzy Associative Classifiers, FDT and Evolutionary Fuzzy Classifiers</p>	<p>The work's goal is to contrast the outcomes produced by the various distributed learning algorithms already covered.</p>	<p>On KDD dataset, each algorithm achieves high accuracy algorithms. The FDTs consistently outperform the FRBCs on the other datasets.</p>	<p>They are reminded that the balanced binary classification datasets for the HIG and SUS are both available. Also KDD dataset</p>
<p>9. Magda M. Madbouly, Eman Abd El Reheem, Shawkat K. Guirguis 2021 [24]</p>	<p>Combine apriori algorithm, Type 2 fuzzy logic, and genetic algorithm</p>	<p>The aim of the proposed method is to maximize accuracy results of redundant rules through pruning. And optimize non-redundant rules using genetic algorithm.</p>	<p>The findings showed that the redundant rules are effectively removed through the suggested algorithm while the typical FARs extracted various rules and different non-redundant rules. As a rule, the suggested method is more effective at eliminating redundant rules in comparison with the existing FARs.</p>	<p>The "Adult" data from the UCI machine learning repository was the dataset that was used. There are 32,659 records in the Adult data set, each with 8 attributes.</p>

4. Literature Analysis

After reviewing these researches, we found the importance of association rule mining techniques in many big data applications in addition to fuzzy logic to find valuable information that can give a great advantageous in decision making process and find a clear relation between different variables. With the aid of several distributed framework (Spark and Map Reduce) big data can handled in a distributed parallel way easily. The major goal is the ability to explore high volume of dataset and elicit a useful knowledge for future action with a reduced time and good accuracy. As seen, many of these works used association rules techniques as apriori and FP- growth to generate strong rules with the assist of fuzzy logic techniques which its role varies in different purposes. To locate the fuzzy rules, several works employed ARM with fuzzy logic for predicting the risk of involvement in detecting diseases like breast cancer with minimum confidence 0.9 and minimum support 0.25. Or by using a unique distributed discretizer and the well-known FP- Growth algorithm to create fuzzy partitions of characteristics. Lastly, these set were pruned. With the use of six very large datasets with up to 11,000,000 cases, they experimented and thoroughly analyzed the outcomes. Concentrating on model complexity, computation time, and scalability. In order to the purpose of preprocessing and analyzing big data utilizing of a hybrid method of fuzzy logic and MapReduce was developed. They presented of MapReduce consist of four layers (as new version). The result showed the efficiency of the approach through Sensitivity = 80%, Specificity = 86% and F-measure= 2.5. By using of K mean algorithm with the using of mode and fuzzy Gaussian functions and compared them to decide the most appropriate one to clustering big data. With big data sets; Covtype, Covtype-2, Poker, and Poker-2. Efficient results were obtained 75.6% accuracy and more execution time in clustering big data. To preserving the data anonymity in big data environment, new association rules hiding technique was proposed with fuzzy logic efficiently. The model verified through experiments on two datasets. Brijs dataset include market basket data, collected from 1999-2000, with 88.162 transactions and 16.469 products. Clue web dataset with one billion web pages in ten languages, subset was used 53,268,952. Judging by metrics like lost, ghost and false rules, with ghost rules in proposed framework are equal in HCSRIL. Finally due to redundant problems of the association rule algorithms making them inefficient and understandable rules to the end users. For these motives, the authors' uses type 2 fuzzy logic Association Rules techniques (T2FARM) to pruning the redundant rules with adapting genetic algorithm to enhance the non-redundant rules. The results shows high accuracy with (98.5) fitness function. However it taking high execution time term as a result of T2FLS computation plus GA takes long time in the improving of rules process. As well as many applications of fuzzy approaches has been conducted recently especially with the appearance of COVID19 disease, to improve diagnosis and minimize misdiagnosis errors.

5. Conclusion

We are living in the era of big data, in which data are produced rapidly and in large quantities from several sources. With algorithms that could discover the correlation between rules and items, ARM is a key DM method. Also fuzzy logic techniques playing a main role in big data processing and reduce the number of generated association rules and helps in enhancing the pruning step to the redundant rules and acts various kinds of uncertainties in the data. We also conclude that:

1. The existing association rule algorithms are improper, inefficient to work with big data and limited with high dimensional datasets.
2. After that, we discovered that the open-source and reliable UCI Machine Learning Repository would be using the medical dataset p53 Mutants. The dataset was chosen due to its huge scale and high dimensionality, which are advantageous scalability properties. The original dataset has 31,420 occurrences and 5,409 features totaling 1.3 GB in size. An active or inactive classification serves as the category feature. Inactive denotes a malignant tumor, while active denotes transcriptionally competent. Utilizing several metrics, such as calculation time and accuracy, to evaluate the effectiveness of the suggested study. Feature selection as well aim to select a few subset features from original features to achieve best results. As seen metaheuristic algorithms are recently used for feature selection as an optimization problem. Such as swarm intelligence (PSO, CS, etc.). Also it has seen data conversion from unstructured form to structure one using hadoop ecosystem tools like: MapReduce and Hbase.
3. Preprocessing steps are also will be implemented mainly via data analysis, feature selection, cleaning of noise and error data to meet requirement of the model which affected by the quality of data. Whereas missing values is filled and delete errors. Also outliers are detected which are so far from normal data to establish a robust model with many factors including: accuracy, scalability and computation time.
4. By a combination of association rule mining and agility fuzzy logic techniques to produce a hybrid model on big data environment with the aid of Spark framework using python as programming language will facilitate working to get great results due to It's Simple, free, easy to use, highly compatible, object-oriented, has lots of libraries, and it's widely applicable.
5. Eventually, there are hugely increasing in generating data on a daily basis from heterogeneous sources. Mining these amounts of data is nontrivial task with common features of big data (volume, velocity and veracity). Many of

proposed works were developed in various domains (healthcare, e-commerce etc.) Using different algorithms and techniques run different dataset in various time complexities and accuracy with diverse output results comparing to other methods to get a useful knowledge. Through examining such extracted rules, ARM is an intriguing method for identifying interesting associations between a variety of data points, uncovering patterns, and, for instance, informing a customer's purchasing behavior.

Funding

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

The author declares no conflict of interest

References

- [1] Alexandru, A., Alexandru, C.A., Coardos, D. and Tudora, E. "Big data: concepts, technologies and applications in the public sector. Int J Comput Electr Autom Control Inform," Eng, 10,, 2016, pp.1629-35..
- [2] ElenaRuiza,, JorgeCasillas "Adaptive fuzzy partitions for evolving association rules in big data stream", International Journal of Approximate Reasoning " ,2018, pp.463-486.
- [3] H. Wei, "Study on a data warehouse mining oriented fuzzy association rule mining algorithm." In 2014 Fifth International Conference on Intelligent Systems Design and Engineering Applications," 2014, pp. pp. 935-938. IEEE.
- [4] Dr. Mortadha M. Hamad and Banaz A. Qader, "Data Pre-processing for knowledge discovery" University of Anbar college of computer, " Tikrit journal of pure science, 2013.
- [5] Armando Segatori, Alessio bechini, PietroDucange, and Francesco Marcelloni , "A Distributed Fuzzy Associative Classifier for Big Data".
- [6] Anish jindal, Amint Dua, Neeraj Kumar, Ashok KUMAR Das, Athanasios V.Vasilakos, and Joel J.P.C. Rodrigues, "Providing Healthcare-as-a-Service Using Fuzzy Rule Based Big Data Analytics in Cloud Computing," IEEE Journal of biomedical and Health informatics , 2018.
- [7] S. a. M. Agarwal, "Analogous Examination of Various Machine Learning Algorithms Applied to Big Data," International Conference on Advances in Computing, Communication Control and Networking (ICACCCN, 2018, pp. pp. 121-123.
- [8] Rathee, Sanjay, and Arti Kashyap, "Adaptive-Miner: an efficient distributed association rule mining algorithm on Spark," Journal of Big Data , 5 (2018), pp. 1-17.
- [9] S. Ahn, S. V. Couture, A. Cuzzocrea, K. Dam, G. M. Grasso, C. K. Leung, K. L. McCormick, & B. H. Wodi, "A fuzzy logic based machine learning tool for supporting big data business analytics in complex artificial intelligence environments,," in FUZZ, 2019.
- [10] Leung, Carson K., Jonathan D. Elias, Shael M. Minuk, A. Roy R. de Jesus, and Alfredo Cuzzocrea. , ""An innovative fuzzy logic-based machine learning algorithm for supporting predictive analytics on big transportation data.," IEEE International Confere, In 2020 .
- [11] Sumathi, R., and E. Kirubakaran. , ""Architectural perspective of parallelizing association rule mining.," In IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM-2012), 2012, pp. pp. 437-442.
- [12] Belbachir, Khadidja, and Hafida Belbachir, "The parallelization of algorithm based on partition principle for association rules discovery," In 2012 international conference on multimedia computing and systems, IEEE, 2012, pp. pp. 1-6.
- [13] Kowsigan, M., A. Christy Jebamalar, S. Shobika, R. Roshini, and A. Saravanan, "Heart disease prediction by analysing various parameters using fuzzy logic," Pakistan Journal of Biotechnology , 2017, pp. 14, no. 2 (2017): 157-161..
- [14] Wang, Liang, David Wai-Lok Cheung, Reynold Cheng, Sau Dan Lee, and Xuan S. Yang. , "Efficient mining of frequent item sets on large uncertain databases,," IEEE Transactions on Knowledge and Data Engineering 24, , 2011, pp. no. 12 (2011): 2170-2183..

- [15] Garg, Kanwal, and Deepak Kumar., "Comparing the performance of frequent pattern mining algorithms.," *International Journal of Computer Applications* , pp. 69, no. 25 (2013).
- [16] Tayal, Kshitij, and Vadlamani Ravi, "Fuzzy association rule mining using binary particle swarm optimization: Application to cyber fraud analytics," *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, , In 2015 , pp. pp. 1-5.
- [17] Segatori, Armando, Alessio Bechini, Pietro Ducange, and Francesco Marcelloni. , "A distributed fuzzy associative classifier for big data," *IEEE transactions on cybernetics*, pp. 48, no. 9 (2017): 2656-2669.
- [18] Hai Wang, Zeshui Xu, Witold Pedrycz, "An overview on the roles of fuzzy set techniques in big data processing," *Trends, challenges and opportunities, Knowledge-Based Systems, Volume 118, 2017*, pp. Pages 15-30..
- [19] Ikhlas Almukahel, Wael AL-zyadat and Mohamad Alfayomi, " Hybrid Approach Using Fuzzy Logic and MapReduce to Achieve Meaningful Used Big Data," *international Journal of Engineering and Technology*, 2018.
- [20] El-Mandouh, Amira M., et al, "Big Data Clustering Model based on Fuzzy Gaussian," *International Journal of Computer Science and Information Security (IJCSIS)* 16.4 , (2018).
- [21] Pejić Bach, Mirjana, Živko Krstić, Sanja Seljan, and Lejla Turulja, "Text mining for big data analysis in financial sector: A literature review," *Sustainability* 11, no. 5: 1277, (2019).
- [22] Madhu, Sake, Ranjit Midde, Gandikota Ramu, Appawala Jayanthi, Jalari Somasekar, Gajula Ramesh, and Pallela Reddy, "A Secured Framework to Protect Association Rules in the Big Data Environment Using Fuzzy Logic," *Ingénierie des Systèmes d Inf.* 24, no. 5 (231-537), 2019.
- [23] Ducange, Pietro, Michela Fazzolari, and Francesco Marcelloni, "An overview of recent distributed algorithms for learning fuzzy models in Big Data classification," *Journal of Big Data* 7, no. 1: 1-29, (2020).
- [24] Magdam. Madbouly, et.al, "Interval type2 fuzzy logic using genetic algorithm to reduce redundant association rules," *Journal Of Theoretical And Applied Information Technology*, 2021.