Republic of Iraq
Ministry of Higher Education and Scientific Research
University of Babylon
College of Information Technology
Department of Software

Enhancement of Privacy Preserving Association Rules Based on Compression and coding Techniques

A Dissertation

Submitted to the Council of the College of Information Technology –

University of Babylon in Partial Fulfillment of the Requirements for the

Degree of Doctorate of Philosophy in Information Technology / Software

By

WAHEED ABDUL-KADHIM SALMAN

Supervised by

Prof. Dr. Eng. Sattar Bader Sadkhan

2021 A.D. **1443 A.H.**

Supervisor Certification

I certify that the dissertation entitled "**Enhancement of Privacy Preserving Association Rules Based on Compression and coding Techniques**" " has been prepared by (**Waheed Abdul-Kadhim Salman**) under my supervision at the department of Software/ College of Information Technology/ University of Babylon as partial fulfillment of the requirements of the degree of Doctor of Philosophy in Information Technology - Software.

Signature:

Supervisor Name: **Dr. Eng. Sattar Bader Sadkhan**

Title: Professor

Date: / /2021

The Head of the Department Certification

In view of the available recommendations, I forward the dissertation entitled "**Enhancement of Privacy Preserving Association Rules Based on Compression and coding Techniques**" for debate by the examination committee.

Signature:

Name: **Dr. Ahmed Saleem Abbas**

Title: Assist. Prof.

Head of Software Department

Date: / /2021

Abstract

With the rapid development of information technologies and data mining, several researchers have tended to solve a central issue related to data mining, which is the privacy preserving data mining.

smart environments data mining that depends on sensors is considered one of the most important and recent types of data mining in recent times.

Association rules mining (ARM) identify what is hidden from the correlations relationships among items and attributes of any environment. These information is considered sensitive, disclose this important knowledge to the public or adversaries is central problem.

The technology of privacy-preserving data mining (PPDM) is an important and modern technology, especially for many applications. PPDM is a mixture of two important specialties: knowledge discovery and data security. PPDM studied extensively with the central, undistributed environments, especially with static data, where there is no time factor and even in the case of the presence of the time factor within the dataset, it is will neglect during the steps and results of mining.

As for our proposed system, it is designed to deal with data streams from sensors in distributed smart environments, where the time factor is taken into account and included in the main steps of the mining Algorithm.

The proposed system includes a data mining Algorithm that works on all sites at the same time. After obtaining the mining results (association rules accompanied by association times), two Algorithms are used to compress and encode the mining results in the different distributed sites (an Algorithm in each site).

The experimental results showed that the mining Algorithm reduces the required time by 1/4 of the time required in the case of using the standard mining Algorithm (Apriori Algorithm), as well as reduces the required storage space.

In addition, and most importantly, a new term appeared in this dissertation that was not known previously, which is the association times that converts the results of data streams mining from the estimation results to the exact and specified at specific times.

As for the final results in the site responsible for calculating the association rules for all sites, the association rules were obtained completely (100%) without losing any association rules or the appearance of fake association rules.

List of Contents

List of Abbreviations

Abbreviations	Meaning of Abbreviations
ARM	Association rule mining
AES	Advanced Encryption Standard
ARAND Algorithm	Association Rule with logical AND operation Algorithm
Ass.from	Association from
Ass.to	Association to
Ass. time	Association time
C	Candidates
CDA	Count Distribution Algorithm
DB	Database
DFIM	Distributed frequent itemset mining
DFSM	Distributed frequent sequence mining
DC	Distributed clustering
DDM	Distributed Data Mining
DM	Data Mining
DARM	Distributed Association Rule Mining
EPPAR-CC	Enhancement of Privacy Preserving Association Rules Based on Compression and coding Techniques
ETL	Extact transform load
EClaT Algorithm	Equivalence Class Transformation Algorithm
F	Frequency
FDM Algorithm	Fast Distributed Mining Algorithm
FP-Growth	Frequent Pattern Growth
FP-Trees	Frequent Pattern Trees
GARE	Global Association Rules Extraction
IoT	Internet of Things
min_conf	Minimum confidence
min_s u p	Minimum support
ODAM Algorithm	Optimized Distributed Association Mining Algorithm
PPDM	Privacy preserving data mining
PPDDM	Privacy preserving of distributed data mining
PPARM	Privacy Preserving Association Rule Mining
PPAR-CC	Privacy Preserving Association Rules based on Compression and Cryptography
RSA	(Rivest–Shamir–Adleman)
SARP-CC	Sensitive Association Rules Protection based on Compressing and Cryptography
SDT-B-ARM	Data Stream Time Based Association Rules Mining
Sup-count	Support count
TDB	Transaction database
U-AR	Unification of association rules
WFPPM	Weighted Fuzzy Privacy Preserving Mining

List of Figures

List of Algorithms

List of Tables

Chapter one
General Introduction

1.1 Introduction

In any environment, data are extremely important part. The role of data mining comes as a result of the massive increase in the amount of data produced by all environments. several techniques like classification, clustering, association rule mining and regression are considered within Data mining tasks. All data mining tasks have the same main steps to extract the knowledge.

Different association rule mining techniques is used to extract important correlation among important data items from large datasets. In order to make better strategic decisions for any environment it is important to share this information in distributed environments. This information contains some sensitive knowledge, like environments based on sensors in all disciplines. During the sharing of this information, sensitive information may leak out to unauthorized persons. This is a major problem, so it is necessary for the information to be extracted and hidden in modern methods to preserve the privacy of these environments and their data [1].

Many applications in real world produce big data. when adding timestamps to the big data called data streams such as sensor data streams. Data stream differ from normal static data that store in different storage structures like warehouses or databases. The most important feature of data in sensor streaming data is that it is dynamic, continuous and change through time.

The privacy preserving data mining is one of the modern techniques that appeared recently as a result of the urgent need for it by data owners in sensitive environment. It performs the process of data mining in different data environments such as centralized and distributed data in some safe methods to preserve the privacy of environments and their data, especially when sharing information. Heuristic based,

Border based, exact based, Reconstruction based and cryptographic based approaches are Some of privacy preserving techniques [2].

In a multi-party computation, cryptography based approaches is often used. Where in several geographical locations, the data is distributed. In many cases, data owners want to share their information with the intended parties without this sensitive information leaking to other unauthorized parties. Horizontal partition distributed data and vertical partition distributed data are the main categories of Cryptographic Based approaches in terms the distribution. Depending on the topology and connection costs standards, mining of association rules is more efficient in vertically distributed data [3].

1.2 Motivation

These days, with the tremendous progress in technologies, especially information technologies, in many times Technology has become very useful, but dangerous at the same time. One of these useful and dangerous techniques is the data mining technique, because it will reveal what is hidden, non-obvious and also considered sensitive from the important knowledge within the data of any environment.

Sharing this sensitive knowledge without taking into consideration the privacy issue will definitely leak this knowledge that is important and sensitive to the public or competitors. This thing is certain to not please the owners of this data or environment. Therefore, data must be managed in a professional manner when shared to preserves its privacy, especially sensitive knowledge. Therefore, technologies recently appeared known as privacy preserving data mining in order to conduct the mining process, taking into account preserving the privacy and confidentiality of the data belonging to the institutions.

In order to assure data owners that data privacy is protected by sober techniques. several researchers motivate to develop modern Algorithms to satisfying need the privacy during data mining process and sharing the knowledge [4].

1.3 Related Work

In this section, we will present in briefly the related works of "privacy preserving association rules mining" of several researchers in recent years: -

1- **In 2015**, Yousra Abdul Alsahib S. Aldeen, Mazleena Salleh and Mohammad Abdur Razzaque, “A comprehensive review on privacy preserving data mining”, This article provides overview on new perspective and systematic interpretation of a list published literatures via their meticulous organization in subcategories. The fundamental notions of the existing privacy preserving data mining methods, their merits, and shortcomings are presented. The current privacy preserving data mining techniques are classified based on distortion, association rule, hide association rule, taxonomy, clustering, associative classification, outsourced data mining, distributed, and k-anonymity, where their noTable advantages and disadvantages are emphasized [5].

2- **In 2016**, Rana Saad Mohammed, Enas Mohammed Hussien, Jinan Redha Mutter, “A novel technique of Privacy Preserving Association Rule Mining” the proposed approach offers new method to maintain the privacy and confidentiality of association rules relying on using stochastic standard map to hide sensitive rules [6].

3- **In 2016**, Masooda Modak and Rizwana Shaikh “Privacy Preserving Distributed Association Rule Hiding Using Concept Hierarchy” offers approach of privacy preserving association rule mining over distributed environment to extracting global association rules. Through data distributed horizontally and vertically, sensitive association rules are securely extracted. In order to mask sensitive rules, the concept of hierarchy has been used [7].

4- **In 2017**, Meenakshi Bansal, Dinesh Grover, Dhiraj Sharma, "Sensitivity Association Rule Mining using Weighted based Fuzzy logic", The proposed method aims to extract all association rules that are considered sensitive by use WFPPM (Weighted Fuzzy Privacy Preserving Mining). Through calculating the weights of the rules, WFPPM completely extract the sensitive rules [1].

5- **In 2018,** Naadiya Khuda Bux, Mingming Lu, Jianxin Wang, Saajid Hussain and Yazan Aljeroudi, " Efficient Association Rules Hiding Using Genetic Algorithms", The proposed Algorithm presents new approach of privacy preserving data mining. It introduced simple genetic encoding of sensitive association rules hiding to protect them. It tries reduce non-sensitive association rules through offers recursive computation [8].

6- **In 2018**, RICARDO MENDES AND JOAO P. VILELA, "Privacy-Preserving Data Mining: Methods, Metrics, and Applications" This paper surveys the most relevant PPDM techniques from the literature and the metrics used to evaluate such techniques and presents typical applications of PPDM methods in relevant fields. Furthermore, the current challenges and open issues in PPDM are discussed [9].

7- **In 2019**, Surendra H and Dr. Mohan H S, "Preserving Privacy of Sensitive Itemsets using Controlled Perturbation of Closed Itemsets" The proposed technique introduces an improved approach by distorting randomly to support of important itemset which is considered sensitive itemsets in the datasets, and the improved approach attempts not to affect relationships among other itemsets [10].

8- **In 2019**, Bharath K. Samanthula, Salha Albehairi and Boxiang Dong, " A Privacy-Preserving Framework for Collaborative Association Rule Mining in Cloud", The proposed technique presents a new approach of privacy

preserving association rule mining, which deals with outsourcing association rules mining problem in federated cloud environment with privacy-preserving manner. The proposed method maintains the security and privacy of sensitive information [11].

9- **In 2020,** Ma´rcio Alencar, Raimundo Barreto, Hora´cio Fernandes, Eduardo Souto and Richard Pazzi," DARE: A decentralized association rules extraction scheme for embedded data sets in distributed IoT devices ", This article describes a method for mining implicit correlations among the actions of IoT devices through embedded associative analysis. Based on support, confidence, and lift metrics,the proposed method identifies the most relevant correlations between a pair of actions of different IoT devices and suggests the integration between them through hypertext transfer protocol requests [12].

10- **In 2020**, Venkatesh Kumar. M and Dr. C. Lakshmi, " AN EFFICIENT SECURE COMPUTATION FOR PRIVACY PRESERVING DATA MINING IN MULTI PARTY COMPUTATION (MPC) – A REVIEW", presents a detailed review of an efficient secure computation mechanism for boosting the privacy and security of data mining procedure in Multi-Party Computation (MPC) approach. It provides an overview of privacy preserving, data mining and multi-party computation. it highlights the usefulness and significance of secure computation techniques for privacy preserving in MPC [13].

11- **In 2021**, Nikunj Domadiya and Udai Pratap Rao, "Privacy Preserving Association Rule Mining on Distributed Healthcare Data: COVID-19 and Breast Cancer Case Study", This research introduced proposed privacy preserving distributed association rule mining scheme with insecure communication channels. They used the concept of an elliptic curve-based

paillier cryptosystem to achieve privacy, authenticity, and integrity and observed some security vulnerabilities in their privacy preserving association rule mining scheme when implemented with insecure communication channels. on complexities with better securities. A case study on the efectiveness of the proposed approach in combating COVID-19 coronavirus and Breast Cancer is also discussed [14].

1.4 Problem Statement and Challenges

Certainly when designing our proposed system "Privacy Preserving Association Rules Based on Compression and coding Techniques " over sensor data streams of distributed environments, we will cope several challenges as follow: -

1- Association Rules Mining in sensor data streams is definitely more difficult than association Rules Mining in static data, because due to dynamic properties of sensor data streams.

2- Association Rules Mining in a distributed environment certainly adds some complexity to the proposed system.

3- Sharing huge data over any network cause time consuming, loosening some data, network congestion, large storage space required.

4- Sharing knowledge over network leads Privacy preserving of environment is compromised and sensitive information is leaked.

1.5 Objectives

1- The aim of this dissertation is to build a proposed distributed system for sensor data streams mining data in secure method.

2- Our system also aims to share knowledge (Association Rules) among sites in a secure manner.

3- Our system consists two phases, the phase one consists of several distributed sites, where in each site a proposed mining algorithm is used to

securely mine sensor data. As for the phase two, it consists of one site called the controller site, and it is responsible for extracting knowledge of the distributed system as a whole in a secure manner.

4- In the phase one, the proposed sensors data stream mining algorithm is used that deals with time efficiently and includes it in the main steps of the algorithm. In addition to that, the proposed algorithm extracts the knowledge through only one scan on the sensor data in order to reduce the execution time and storage space, after that two algorithms are used to compress and encode the extracted knowledge and send it from each site to the controller site.

5- In the phase two, this compressed knowledge is received from each site, decoding and decompression operations are performed on it, and then the special knowledge of the distributed system as a whole is extracted in a secure and accurate manner without any false negative [15].

1.6 Problem Solution and Contributions

Designing proposed system for privacy preserving association rules mining in distributed environments consists the following: -

1- 1- In this work, we conduct mining process on sensor data streams by the proposed method through using one scan only on sensors data in order to reduce execution time & storage space, the result will be extracting important correlation with **association times**.

2- After obtaining the resulting knowledge in each site, we can employ **two new proposed techniques** to compress and coding sensitive knowledge, before sending it.

3- In each site over distributed system, we use one of two proposed new approaches to protect the knowledge (association rules) from unauthorized persons.
4- Sharing only small knowledge (very small and incomprehensible text for each site) over network instead of huge data this leads **reduce time required**, and **reduce storage space required**. In addition, to maintain the **privacy** and **confidentiality** of the extracted knowledge from unauthorized persons.
5- Eliminate the **false negative problem**, because every association rule within the global association rule exists within association rules for one or more sites.

1.7 Dissertation Organization

The dissertation consists of five chapters including chapter one. They are as follows:

1- Chapter Two "Theoretical Background " this chapter presents concepts and fundamentals of Static data versus streaming data, Privacy preserving data mining over data streams, Privacy preserving data mining over Centralized and distributed data, Privacy preserving association rules mining, Association rules Algorithms, Association rules mining in smart environments, Privacy preserving association rules mining over distributed environments, base64 coding Algorithm, Huffman coding approach.

2- Chapter Three "*A Proposed system (EPPAR-Cc)*: Enhancement of Privacy Preserving Association Rules Based on Compression and coding Techniques. This chapter presents the details of the proposed system are explained, which consists of five main algorithms, Data Stream Time based Association Rules Mining (DST-B-ARM) Algorithm, Sensitive Association Rules Protection based on Compressing and Coding (SARP-CC) Algorithm, Privacy Preserving Association Rules based on Compressing and Coding (PPAR-CC) Algorithm, Unification Association Rules (U-AR) Algorithm, Global Association Rules Extraction (GAR-E) Algorithm.

3- Chapter Four "*Discussion and Experiment Results"* In this chapter some experimental work is implemented on the Proposed system (EPPAR-Cc), and is compared with results with similar related approaches.

4- Chapter Five "*Conclusions and Suggestions for Future Work"*, This chapter explains the conclusions that were clarified after conducting practical experiments for the proposed system (EPPAR-Cc), Also the chapter gives the suggestions for future work.

Chapter two
"*Theoretical Background* "

2.1 introduction

Day after day, the world depends more and more on data and information technology, especially in the twenty-first century. From the diverse areas of life generate very huge data transactions and in various social, commercial, financial, military, medical and other specialties to include data, in video, digital images, sound and sensors [16][17]. The increase in collecting, storing and analyzing data for institutions or individuals causes a challenge to protect privacy and the level of the challenge is according to the nature of the data [18]. Data analysis techniques such as knowledge discovery techniques analyze data collected from various sources to produce valuable knowledge. One of the most important knowledge discovery techniques is data mining which means mining information from huge records of datasets [19].

In the current decade, data mining has become an urgent necessity, and even in the coming decades it is expected that it will be one of the areas of research that is important to develop for researchers [20]. In order to obtain useful information from a large dataset, data mining has become a necessity and is required [21]. The process of extracting useful and valuable knowledge from massive data is called data mining process. In order to discover valuable and important knowledge of a large collection of data, various data mining techniques can be used [22]. privacy preserving data mining includes two main parts, the first is the data mining process and the second is the preservation of the privacy of the mining results as explained in Figure (2-1).Most of the time the knowledge extracted from the data mining process is sensitive, so the leakage of this sensitive extracted information to the public or competitors during the mining process is a big problem that must be solved.

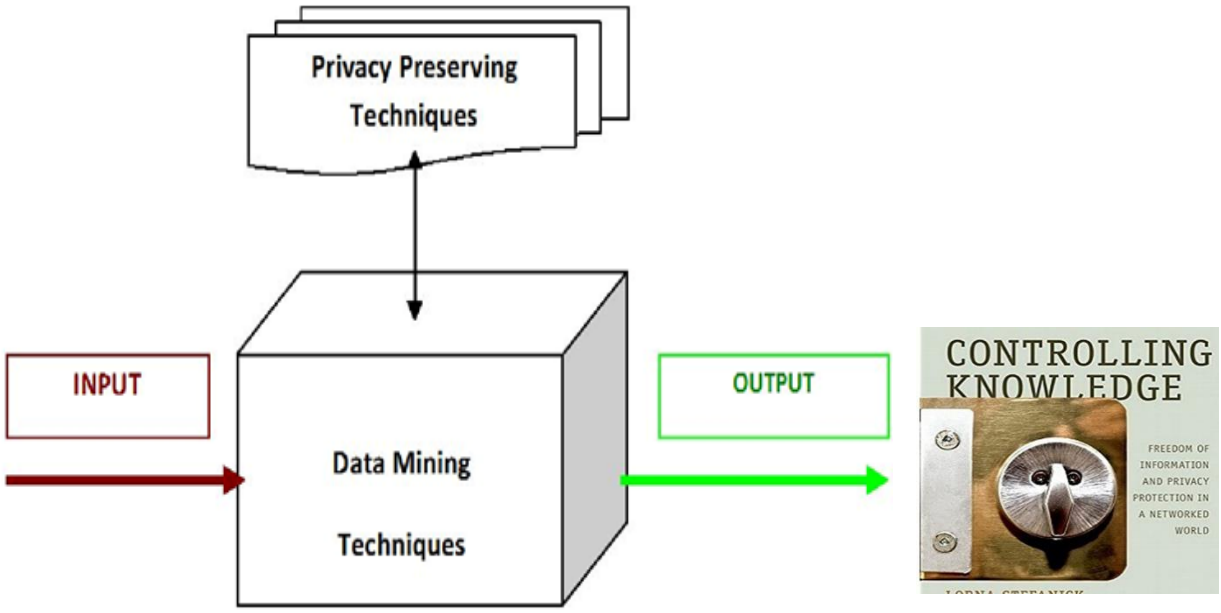


Figure (2-1) privacy preserving data mining [4]

Maintaining privacy for organizations and individuals has become a major and crucial issue . Therefore, the balance between the mining process and maintaining privacy is very important issue and must be managed professionally [23].

Sharing the information extracted after the data mining process between the institutions for the mutual benefit is very important, but during the sharing and exchange of this important and sensitive information, some sensitive information may leak out to unauthorized people[24]. Therefore, it is the turn of privacy preserving data mining (PPDM) techniques to solve this problem and professionally manage this sharing of information so that it reaches only authorized persons. Perturbation, Secure Sum Computations and Cryptographic based techniques are considered within the main classification of Privacy Preserving Data Mining techniques as explained in Figure (2-2) [25].

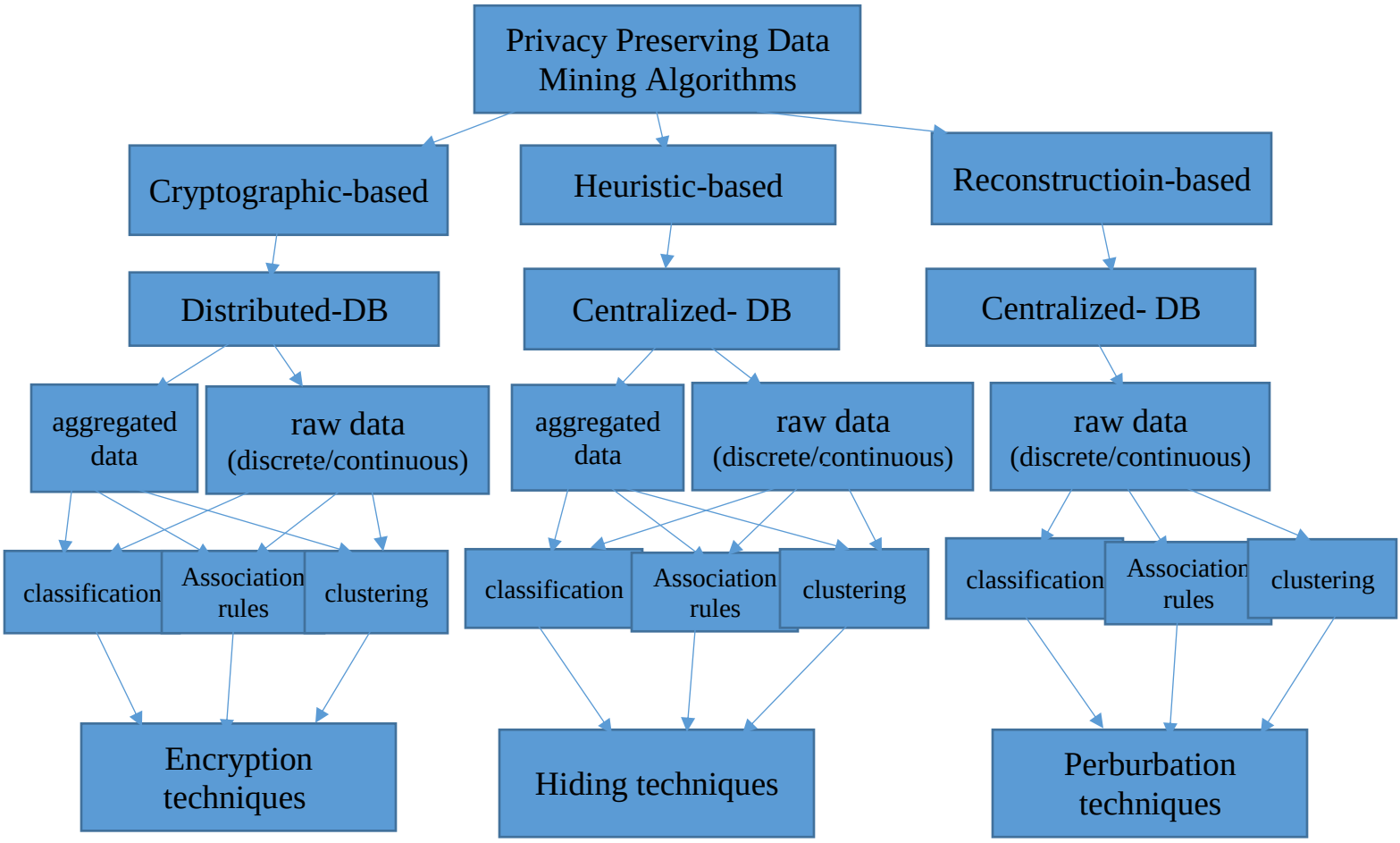


Figure (2-2) Main classification of Privacy Preserving Data Mining [26]

2.2 Static data versus streaming data

The difference between static data and data streams in terms an operational point of view explains that when the data is stored, then analyzed and processed.

Before analysis and processing, static data is stored in a file or database. After the analysis and processing, the results are stored either in a file or records of database to be ready after that for further analysis and processing or it stores in the form of final results that can be used to assist in making decisions and others.

The data is static, if each data observations are recorded and stored before the analysis and processing of data is performed.

The analysis and processing of the streaming data takes place without any storage, ie, only events. Events are analyzed and processed through read-only, without

storage. The results of analysis and processing are stored in a file or records of database. Events are discarded after analysis and processing if data, meaning the data are missed forever in the case of data streams as in sensors data streams in smart environments [27].

There may be the same source for static and streaming data. It may be analyzed with the same analysis tools (data mining), but most of the time the Algorithms differ for analyzing and processing static data and streaming data, because the Algorithms for analyzing and processing static data deal with stored data and not only events as in Algorithms that analyze and process, streaming data. as storage is not necessary. As for the results after performing analysis and processing operations for the static and flowing data, they are stored in the same way in a file or records for a database as explained in Figure (2-3).

Let a simple example of one social media out there is Twitter. Whereas, the logs are analyzed for incoming data from the Twitter application. Whereas, the data analyst wants to categorize tweets by the text content of the tweet for a certain period of time. It can analyze and process static and streaming data from the same source i.e. Twitter application [27].

As the data analyzer stores all the textual content of the tweets for certain period of time within a database. the data is analyzed and processed after the storage process. The results of the analysis and processing are reported and stored in a file or database. This is in the case of static data.

Unlike static data, in streaming data where the textual content of each tweet is analyzed by reading only without storing. Where only the results from the analysis are reported and stored in a file or database After that, all data streams events are neglected.

Finally, and in summary, in the static data, the raw data and results after analysis and processing must be stored. If the data volume is relatively small and the storage space is sufficient, the use of static data is efficient.

But in the case that the volume of data is big and the storage space is limited that it cannot store the volume of data, then the use of data streams is more efficient [27].

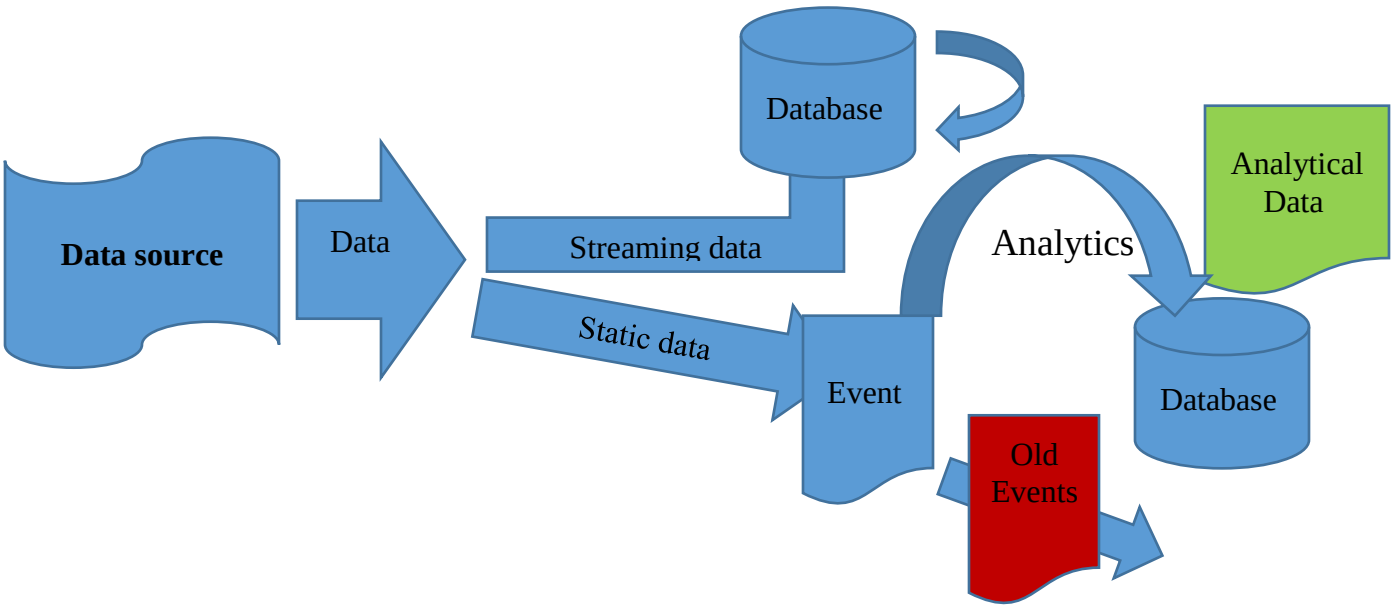


Figure (2-3) static data versus streaming data [27].

2.3 Privacy preserving data mining over data streams

Data mining is one of the most prominent knowledge discovery techniques, which finds reliable and useful knowledge from huge amount of datasets [28]. Recently, a new type of data has started to appear and spread, which is called data streams such as sensor data streams, which originally has its properties that distinguish it from static data. This new type of data includes the time factor, the data continue to flow and changes with time continuously. These are considered important characteristics of data streams [29].

The classic data mining Algorithms that deal with static data are often invalid for handling streaming data.

This is because in most cases classical Algorithms needs to re-scan the database more than once, and this is not possible with the streaming data in addition to other features that add additional complications to dealing with this data, such as the time factor, the continuous change of data, and others. In addition, the techniques for preserving privacy data mining that have been extensively studied in the last period are also designed to deal with static data. As for the streaming data, it varies because it has dynamic properties. Therefore, the privacy preservation Algorithms for data mining are not suitable for dealing with streaming data that has dynamic properties. [30], [31].

In data mining and streaming data mining, the problem of preserving the privacy of sensitive data is a critical issue. Always there should be a balance between utility of data and maintaining privacy [32]. Therefore, the problem of the privacy preserving for data streaming mining is a very big issue. Several parameters lead to the success of privacy-preserving Algorithms for streaming data mining, including performance, accuracy, level of complexity, utility of data and others.

Many privacy preserving data mining techniques that deal with static data are not suitable for dealing with streaming data due to the dynamic properties of the streaming data. For example, Geometric data perturbation techniques that works well with static data but is not suitable for dealing with streaming data and others of technologies.

In some cases, there are privacy preserving data mining techniques that can deal with static data and streaming data, for example privacy preserving data mining techniques based on cryptography, but these techniques have their drawbacks such as computational complexity, relatively long time, relatively large storage area, and others [33].

Data streams have different sources and types, for example, sensor data streams in smart environments, video data streams system, stock market, as well as Internet traffic as explained in Figure (2-4).

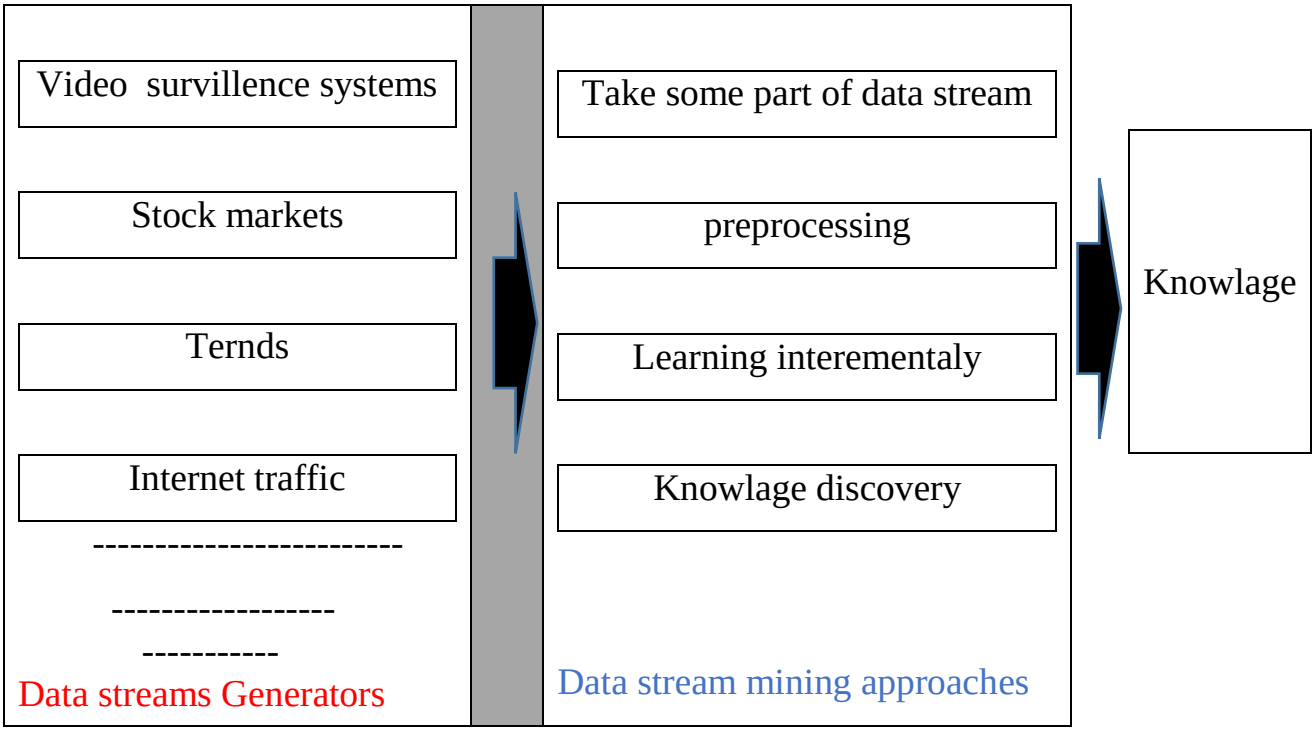


Figure (2-4) Different sources and types of data streams

2.4 Privacy preserving data mining over Centralized and distributed data

Data, by its nature, are either stored, available in one location (centralized data), or distributed across multiple locations (distributed data). In the case of centralized data, data mining and privacy preserving data mining may constitute a challenge in order to preserve the sensitive information contained within the dataset belonging to an organization. In the case of distributed data, data mining and privacy preserving data mining become more difficult and more complicated than privacy preserving data mining in centralized data, as explained in Figure (2-5)[34].

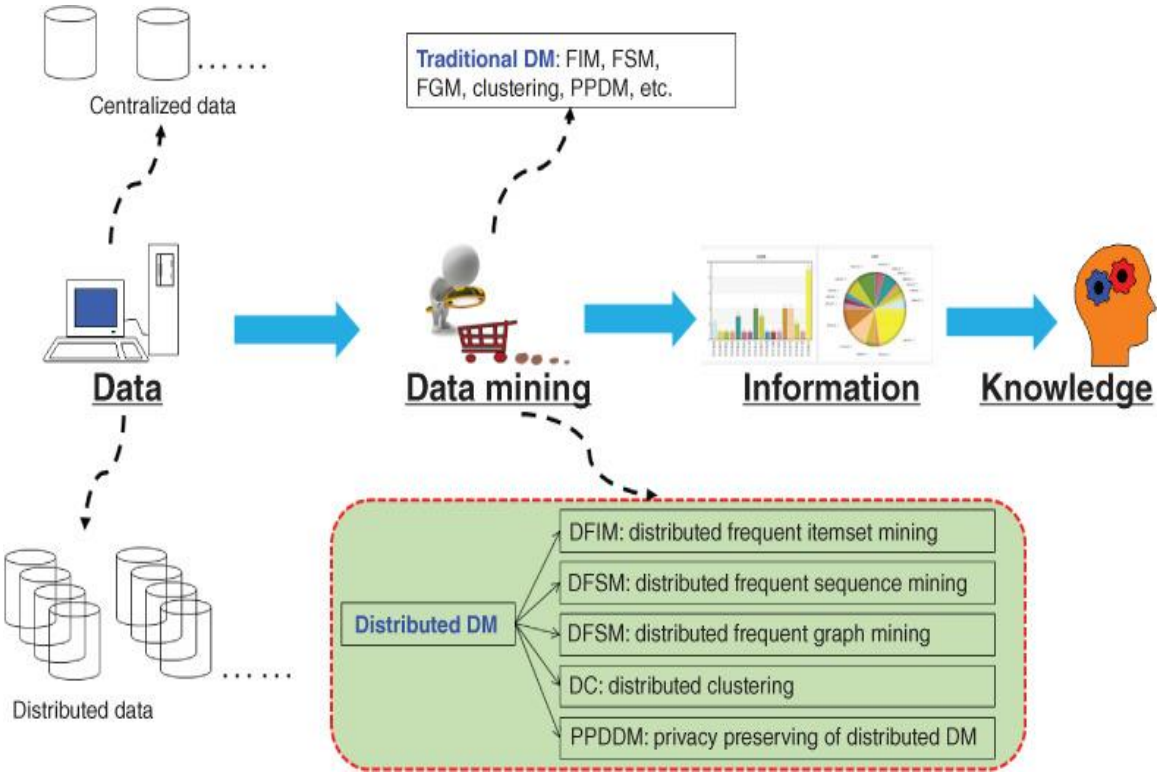


Figure (2-5) Centralized and distributed data mining [35]

Extracting reliable and useful information from multiple huge data sources is data mining aim, it is what data holders need from data mining techniques. But after or during the data mining process, some sensitive information may be leaked intentionally or unintentionally, and this case is considered a major problem for data owners. Therefore, the concepts of privacy and preserving privacy in the field of data mining appeared to solve these problems and preserve the non-leakage of sensitive and important information except to authorized persons only and thus preserve the privacy any environment and their data, especially if the data and information are transferred among multiple parties, so the security and privacy of information becomes more complex.

Data distortion based techniques, clustering based techniques, intersection based techniques, data distribution based techniques and cryptography based techniques are the most important techniques in a field of privacy preserving data mining [34].

In various distributed system environments, the issue of security and privacy becomes more difficult and complex, and the leakage of sensitive information during and after the data mining process has very high probability. To overcome the problem of leakage of sensitive information and violation of the security and privacy of individuals and institutions. Several techniques have been proposed to preserve the privacy of distributed data mining, including randomization based privacy preserving techniques and cryptographic based privacy preserving techniques [36].

In order to exchange sensitive information reliably and securely among multiple parties in distributed data environments and others in terms of analyzing and processing data and ensuring its correctness as well as publishing, the privacy Preserving data mining has emerged as an absolute prerequisite. The data security threat is escalating dramatically in the Internet and causes a great problem to spread sensitive information between multiple parties in a manner that guarantees security and privacy. On the other hand, many researchers have worked and developed new technologies in order to give assurance to data owners that their sensitive information can be shared to authorized or intended persons only in a way that preserves the security and privacy of the shared data [37, 38].

In collaborative data mining, the study and analysis in order to propose new techniques to preserve the privacy of distributed data mining is a critical issue to improve the efficiency and effectiveness of the exchange of sensitive information for static and streaming data within distributed environments that ensure the security and privacy of the information shared [39].

2.5 Privacy preserving association rules mining

The most important methods for discovering knowledge is data mining, through which patterns or important knowledge are extracted from huge amounts of data by means of many diverse data mining techniques. Information mining and security protection are considered important and worried fields, and there is an implicit link

between them, as data mining produces important and sensitive knowledge, and this important knowledge certainly needs protection in order to reach only the intended people only not to public [40].

The techniques through which the link between information security and information mining is legitimized are techniques for privacy preserving data mining (PPDM) that work to extract un-intuitive knowledge from the vast data taking into account preserving the security and privacy of sensitive information extracted in many applications.

Many researchers have made unremitting efforts to develop methods privacy preserving data mining and preserve the security of extracted information [41]. The best system that can be used to extract hidden knowledge within huge data is a data mining system through which patterns and communications between objects are extracted and discovered in an efficient and effective manner [42]. Maintaining security and privacy has become an important issue in the field of mining and disclosure of information, especially after the great increase in sharing this sensitive and accurate information between organizations, governments and different gatherings. The most common and important method that has been extensively studied within the techniques of analyzing and mining data for a wide range of applications is association rules mining [41].

Association rule mining exposes important correlations and patterns among items and features in huge datasets [43]. One of the most efficient data mining techniques is Association rule mining, through which the frequent items and Associative rules are identified by analyzing market basket data for massive amounts of transactions. Association rule mining is essential for organizations to support strategic decision-making for their success as explained in Figure (2-6).

Association rules mining basically means calculating the probability of appearing for the most frequent items according to the appearance of another item in the transactions database.

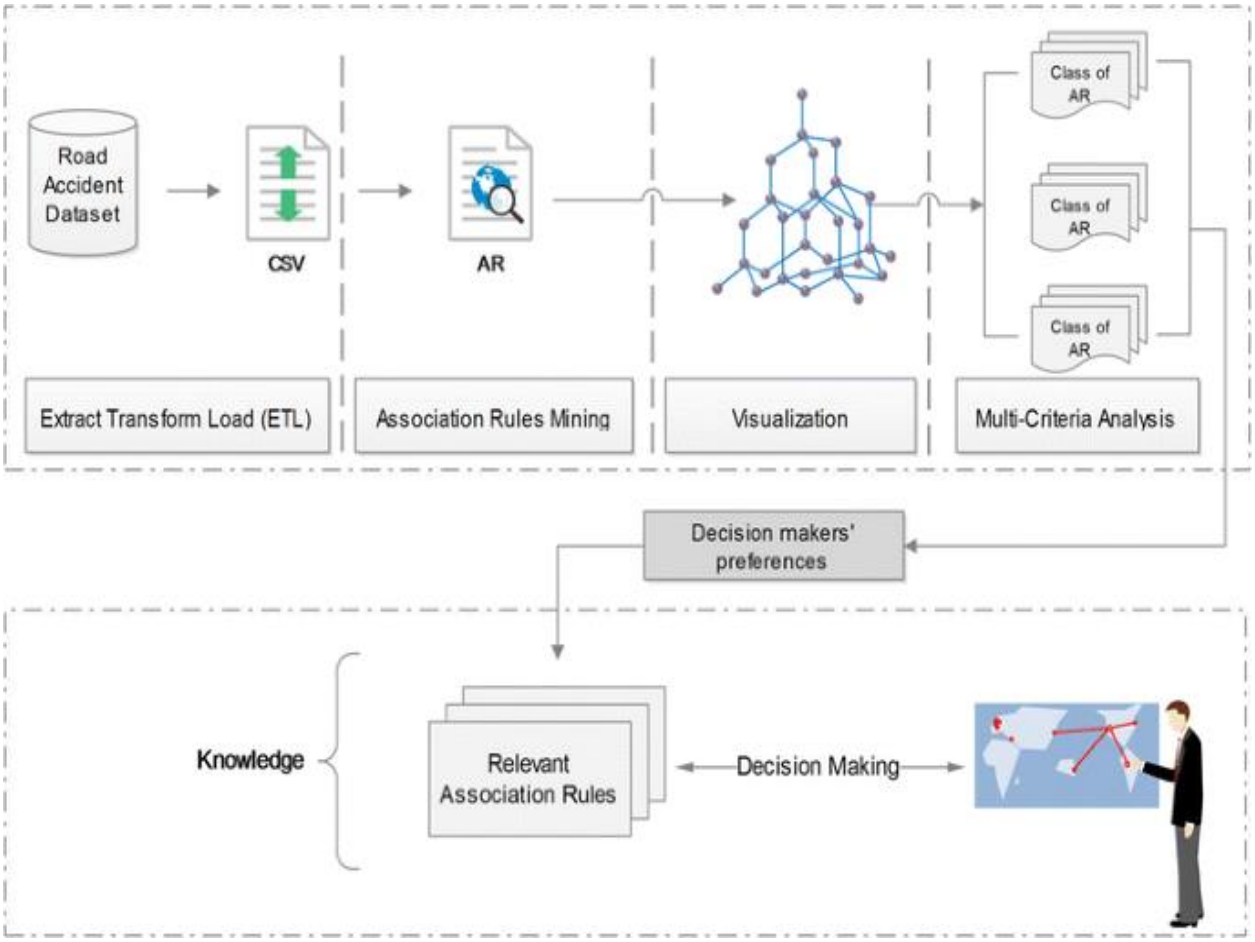


Figure (2-6) important of association rules for decision making [44]

Association rules mining will certainly reveal the important and sensitive patterns and connections hidden within the datasets, and this may lead to revealing the confidentiality and privacy of individuals and institutions. In order to solve the problems of security threats during and after the data mining process, it has been suggested and used Privacy Preserving Data Mining (PPDM) techniques. Likewise, in order to preserve the security and privacy of data in terms of not disclosing patterns and delicate important correlations among the items and features, which in

turn violates the privacy of individuals and organizations, it was suggested and used that were then improved and developed Privacy Preserving Association Rule Mining (PPARM) techniques [45].

privacy-preserving association rules mining is the process of protecting and hiding sensitive knowledge (association rules) using several privacy-preserving techniques as explained in Figure (2-7) [46].

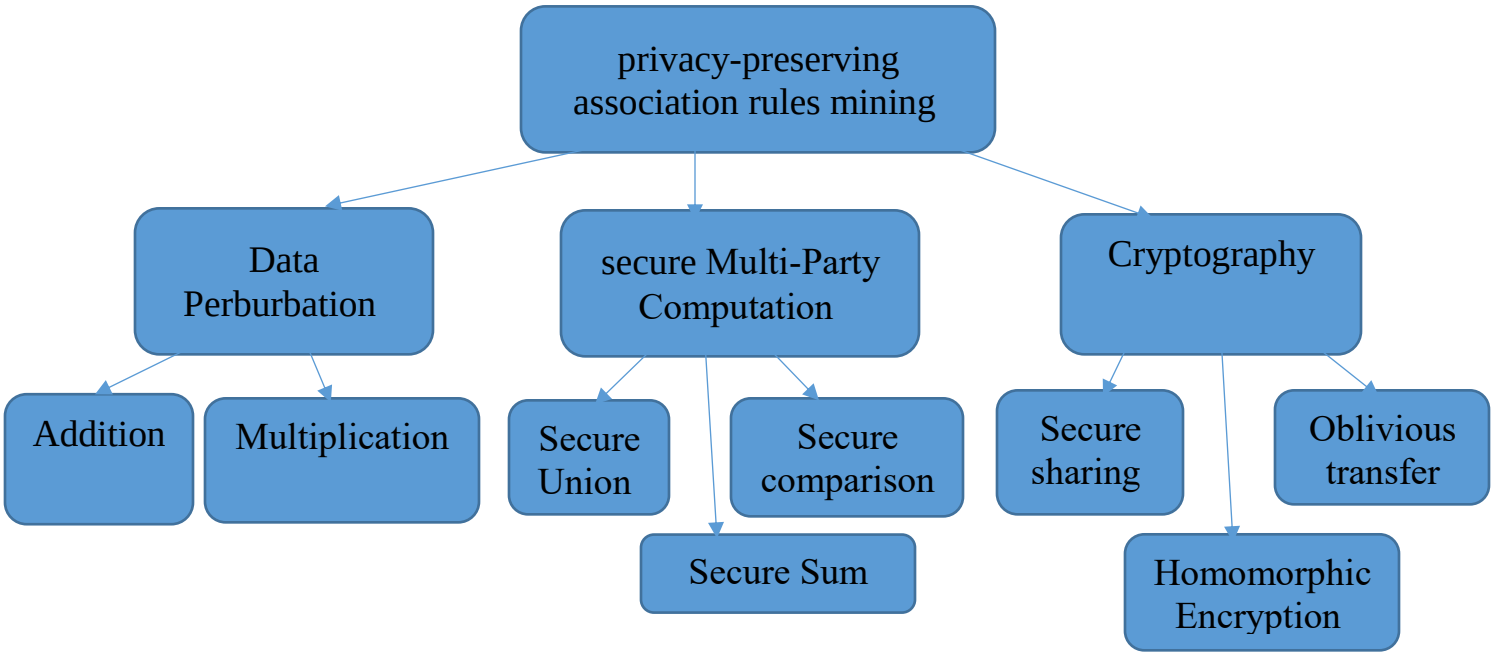


Figure (2-7) Types of privacy-preserving association rules techniques [47]

Recently, privacy-preserving association rules mining Algorithms have been proposed to support data security and it privacy [48].

Research in the field of privacy-preserving association rules mining is considering an open problem as the recently proposed technologies which contain disadvantages and always they need improvement and development in order to reduce these disadvantages.

2.5.1 Association rules mining

Data mining are one of the most important knowledge discovery tools through which interesting and previously unknown patterns are detected from a very large amount

of data. One of the most important data mining techniques, through which interesting and previously unknown patterns and correlations relationships hidden in a huge database are discovered is called Association rule mining [49].

Association rule mining Algorithms mainly include two steps. the first step is mining of frequent itemsets and the second step is generating strong association rules. There are two main parameters of association rules mining are support threshold and confidence threshold as explained in Figure (2-8).

Within real-world gigantic data and over a wide range of applications from all disciplines fast and efficient association rule mining techniques emerge as a valuable approach [50].

In smart business applications as well as in the intelligent manufacturing system, association rule mining techniques is widely used in order to make better smart decisions automatically [51].

In the past years, several efficient association rules mining Algorithms have been proposed to deal with binary or discrete-valued data as inputs [52].

In real world applications, there are many different types of data.

The classic association rules mining Algorithms deals with data that is in the form of transactional databases, meaning the representation of items or features based on whether or not the item or set of items appears in the transaction. association rules are measured in terms of the probability that a set of frequent items will appear relative to the probability of a frequent item appearing.

On the other hand, in other type of applications where the data has a numerical value. In this case, classic association rules mining Algorithms are not suitable for handling this type of data. Therefore, association rule mining Algorithms have been developed to deal with numerical values, and so-called fuzzy association mining Algorithms.

On the other hand, classical association rules Algorithms deal with static data and have no dynamic properties. In the case of dynamic data that is constantly changing with time, another type of Algorithm called online association rule mining has been developed.

The association rules Algorithms have evolved more and more to deal with most other types of data such as unstructured or unorganized data, as well as with text data such as text mining Algorithms for association rules [53].

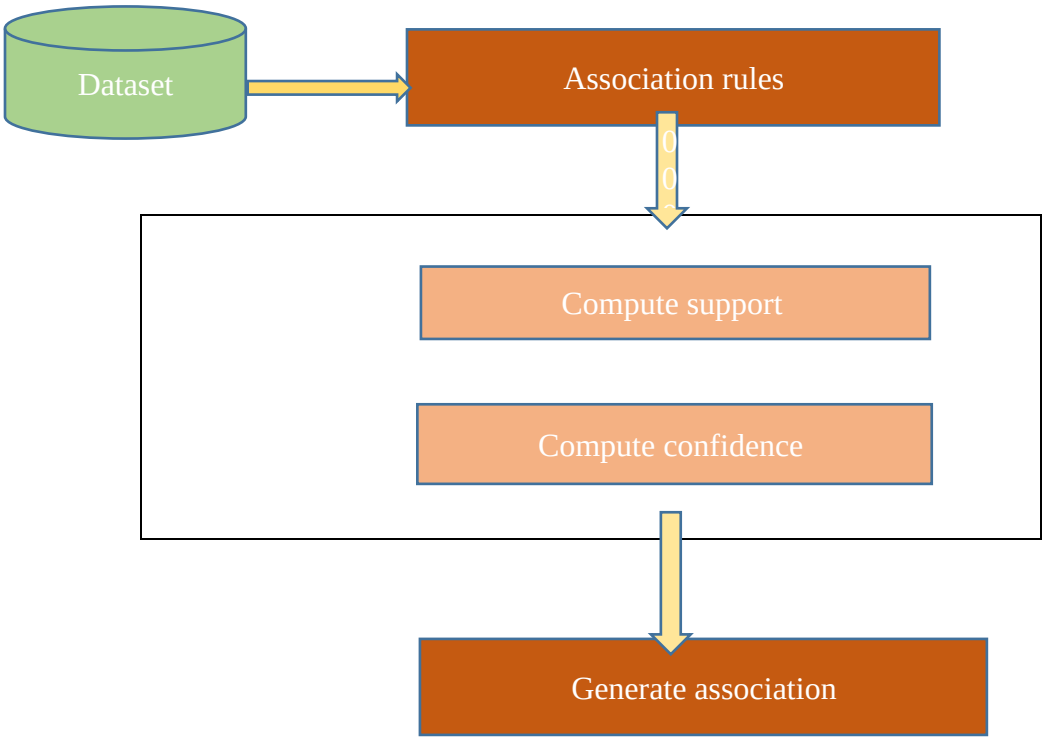


Figure (2-8) steps of association rules generation [50]

2.5.2 Association rules Algorithms

Data mining Algorithms are primarily designed to discover new knowledge hidden within a huge database by using iterative process. In massive databases, the data mining Algorithm may take a relatively long time during the implementation process. association rules mining Algorithms are considered the most important and most popular among data mining and knowledge discovery Algorithms. In massive databases, the association rules mining Algorithms suffer from some shortcomings.

However, several proposed approaches, including one that dealt with incremental data issues, mean this when large amounts of data are added to the database, especially after performing the mining process [54].

Association rule mining Algorithms are basically applied to discover the most important correlations relationships between things. There are many basic Algorithms for Association rules mining, the most important of which is Apriori Algorithm, FP-Growth Algorithm, EClaT Algorithm, ARAND operation Algorithm [55].

1- Apriori Algorithm

Apriori Algorithm (Agrawal 1994) is the most popular Algorithm for extracting frequent itemsets, it is based on scan & repeated re-scan of transaction database.

Steps of the Apriori Algorithm

1. Computing the support for each individual item.
2. Deciding on the support threshold.
3. Selecting the frequent items.
4. Finding the support of the frequent itemsets.
5. Repeat for larger sets.
6. Generate Association Rules and compute confidence.

Where the search technique at the level of k-itemsets is used to extract the level of k+1itemsets and so on. For example, there is simple transaction database which contain different transaction as explained in Table (2-1). At first the transaction database is scanned to produce set of frequent 1-itemsets, each of them checked as in Table (2-2). Each of these must be at least equal to the minimum support threshold. After that, the 2-itemset are formed from the frequent 1-itemsets as in Table (2-3), Then the database is rescanned to obtain frequent 2-itemsets as in Table (2-5). In order to identify all k-itemsets, the entire transaction database must be

scanned until a level is reached at which it is not possible to obtain frequent itemsets where scanning stops [56].

Table (2-1) simple transaction database

TID	List of item_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Compute candidate support count with minimum support count, L1

Scan transaction database for count of each candidate, C1

Table (2-2) support of 1-itemsets

Itemsets	Sup-count
11	6
12	7
13	6
14	2
15	2

Generate C2 candidates from L1, Scan database for count of each candidate

Table (2-3) 2 itemsets Table

Itemsets
11, 12
11, 13
11, 14
11, 15
12, 13
12, 14
12, 15
13, 14

13, 15
14, 15

Table (2-4) support of 2-itemsets

Itemsets	Sup-count
11, 12	4
11, 13	4
11, 14	1
11, 15	2
12, 13	4
12, 14	2
12, 15	2
13, 14	0
13, 15	1
14, 15	0

Compute candidate support count with minimum support count, L2

Table (2-5) frequent of 2-itemsets

Itemsets	Sup-count
11,12	4
11,13	4
11,15	2
12,13	4
12,14	2
12,15	2

Generate C3 candidates from L2, Scan database for count of each candidate

Table (2-6) 3-itemsets Table

Itemsets
11,12,13
11,12,15

Table (2-7) support of 3-itemsets

Itemsets	Sup-count
11,12,13	2
11,12,15	2

Compute candidate support count with minimum support count, L3

Table (2-8) frequent of 3-itemsets

Itemsets	Sup-count
11,12,13	2
11,12,15	2

Algorithm (2-1): Apriori Algorithm [57]

```
// Finds frequents itemset using an iterative level-wise approach based
   on candidate generation.
  Input:
  Database D of transactions;
  min-sup;
  Output: frequent itemsets.
  Begin
    L1 = find frequent 1-itemsets(D);
    For (k=2; Lk - 1 ≠Ø; k++)
    {
        Ck = apriori_gen(Lk-1, min_sup);
        For each transaction t ∈ D //scan D for counts
        {
            Ct = subset(Ck,t); //get the subsets of t that are candidates
            For each candidate c ∈ Ct
                c.count++;
        }
      Lk= {c ∈ Ck|c.count >= min_sup}
    }
   return L =  ∪k Lk
  End.
```

2- FP-Growth Algorithm

Frequent Pattern Growth Algorithm (FP-Growth) (Han et al. 2000) is one of the most important and popular Algorithms of frequent itemsets mining. Unlike Apriori Algorithm, FP-Growth Algorithm works to find frequent itemsets without generating candidate. This leads to a reduction substantially the cost of searching for

the frequent patterns. if the data set is very large, then Frequent Pattern Growth Algorithm is time consuming.

Steps of FP Growth Algorithm

1- Scan DB once, find frequent 1-itemset (single item pattern)

2- Sort frequent items in frequency descending order, f-list.

3- Scan DB again, construct FP-tree.

4- Construct the conditional FP tree in the sequence of reverse order of F - List - generate frequent item set.

FP-Growth constructs tree of frequent patterns to compress frequent items by using a divide-and-conquer technique, where the tree of frequent patterns is divided further to consist Conditional FP-Trees, after that each frequent item separately is mined. As explained in simple example of steps for extract frequent itemsets based on Frequent Pattern Growth Algorithm as in Table (2-9) [56].

From transaction database of Table (2-1)

Table (2-9) Conditional pattern base & Conditional FP-Tree

Items	Conditional pattern base	Conditional FP-Tree	Frequent pattern
15	{(12,11:1), (12,11,13:1)	(12:2,11:2)	(12,15:2), (11,15:2), (12,11,15:2)
14	{(12,11:1), (12:1)}	(12:2)	(12,14:2)
13	{(12,11:2), (12:2), (11:2)}	(12:4,11:2), (11:2)	(12,13:4),(11,13:4),(12,11,13:2)
11	{(12:4)}	(12:4)	(12,11:4)

Algorithm (2-2): FP-Growth Algorithm [58]

Input: A database DB, represented by FP-tree constructed according to Algorithm 1, and a minimum support threshold?

Output: The complete set of frequent patterns.

Method: call FP-growth (FP-tree, null).

Procedure FP-growth (Tree, a)

Begin

{

1- if Tree contains a single prefix path then {// Mining single prefix-path FP-tree

2- let P be the single prefix-path part of Tree;

3- let Q be the multipath part with the top branching node replaced by a null root;

4- for each combination (denoted as ß) of the nodes in the path P do

5- generate pattern ß ∪ a with support = minimum support of nodes in ß;

6- let freq pattern set(P) be the set of patterns so generated;

}

7- else let Q Be Tree;

8- for each item ai in Q do {// Mining multipath FP-tree

9- generate pattern ß = ai ∪ a with support = ai. support;

10- construct ß's conditional pattern-base and then ß's conditional FP-tree Tree ß;

11- if Tree ß ≠ Ø then

12- call FP-growth (Tree ß , ß);

13- let freq pattern set(Q) be the set of patterns so generated;

}

14- return (freq pattern set(P) ∪ freq pattern set(Q) ∪ (freq pattern set(P) × freq pattern set(Q)))

}

END

3- EClaT Algorithm

Equivalence Class Transformation Algorithm (EClaT) (Zaki 2000), It is one of the well-known Algorithms that efficiently mine the frequent itemsets, depending on the use of the vertical data format, as all transactions that contain a specific element are collected in the same record [56].

Steps of EClaT Algorithm

1- List the Transaction ID (TID) set of each product.

2- Filter with minimum support.

3- Compute the Transaction ID set of each product pair.

4- Filter out the pairs that do not reach minimum support.

5- Continue as long as you can make new pairs above support.

In the beginning This Algorithm (EClaT) performs a single scanning process for the database through which the representation of the data is converted from the horizontal formula to the vertical formula. Where frequent 1-itemsets are produced during the first scan as in Table (2-10). Whereas frequent 2-itemsets are extracted by performing an intersection operation for frequent 1-itemsets as in Table (2-11). This means that frequent k+1itemsets are extracted from the intersection of frequent k-itemsets. This process is repeated until all levels of frequent itemsets are found as in Table (2-12).

This Algorithm (EClaT Algorithm) distinguished from other Algorithms for finding frequent itemsets by performing only scan the database once to find frequent 1-itemsets. As for finding the rest levels of frequent itemset it just performs intersection operation of the k-itemsets with one another [56].

From transaction database of Table (2-1)

Table (2-10) transaction database in vertical data format (1-itemset)

Item sets	Tid-set
11	{T100, T400, T500, T700, T800, T900}
12	{T100, T200, T300, T400, T600, T800, T900}
13	{T300, T500, T600, T700, T800, T900}
14	{T200, T400}
15	{T100, T800}

Table (2-11) 2-itemsets in vertical data format

Item sets	Tid-set
11,12	{T100, T400, T800, T900}
11,13	{T500, T700, T800, T900}
11,14	{T400}
11,15	{T100, T800}
12,13	{T300, T600, T800, T900}
12,14	{T200, T400}
12,15	{T100, T800}
13,15	{T800}

Table (2-12) 3-itemsets in vertical data format

Item sets	Tid-set
11, 12, 13	{T800, T900}
11, 12, 15	{T100, T800}

Algorithm (2-3): Eclat_growth-Algorithm [59]

Input: (Database: D, Min_sup: MS)

Output: All FIs L1.

begin

Define: VM : Vertical Matrix, PT: Two-dimensional Pattern Tree, FIs:

Frequent VM = CreateVMfromDatabase(Database: D) L2.

PT= CreateNullPatternTree L3.

for i = 1 to length(VM) do L4.

if length(VM[i].TID_sets) >= MS then L5.

AddItemsettoPatternTree(VM[i], PT, MS); L6.

end //

end

for i = 1 to length(VM) L7.

FIs=GetAllFrequentItemsetsfromPatternTree(Two-dimensional Pattern Tree: PT)

End

4- ARAND operation Algorithm

Association Rule with logical AND operation (ARAND**)** Algorithm (Emad Kadum Jabbar, 2005) [60].

This Algorithm works according to AND operation logic where it begins by encoding the transaction database in binary encoding where 1 is for the item that appears in the transaction and 0 for the item that does not appear in the transaction, as in the following Table (2-13) & Table (2-14): -

Table (2-13) Example of TDB

TID	items
T100	A,B,E
T200	B,D
T300	B,C
T400	A,B,D
T500	A,C
T600	B,C
T700	A,C
T800	A,B,C,E
T900	A,B,C

Table (2-14) binary coding of TDB (1-itemsets)

TID	A	B	C	D	E
T100	1	1	0	0	1
T200	0	1	0	1	0
T300	0	1	1	0	0
T400	1	1	0	1	0
T500	1	0	1	0	0
T600	0	1	1	0	0
T700	1	0	1	0	0
T800	1	1	1	0	1
T900	1	1	1	0	0

After that, each column in the Table (2-14) is checked. In the case that the number of 1s is less than the minimum support threshold, then this column is neglected and proceed to the next step.

Where we consider Table (2-14), Table of 1-itemset, from this Table we construct Table of 2-itemset by using AND logic operation between columns of Table 1-itemsets.

For example, from column A which is (100110111) and column B which is (111101011) we will construct column AB in Table of 2-itemset by applying AND

logic operation on column A and column B. the result will be column AB which is (100100011) of Table 2-itemsets as in Table (2-15) [60].

After that, each column in the Table (2-15) is checked. In the case that the number of 1s is less than the minimum support threshold, then this column is neglected as in Table (2-15) and proceed to the next step [60].

Table (2-15) Table of 2-itemsets

TID	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
T100	1	0	0	1	0	0	1	0	0	0
T200	0	0	0	0	0	1	0	0	0	0
T300	0	0	0	0	1	0	0	0	0	0
T400	1	0	1	0	0	1	0	0	0	0
T500	0	1	0	0	0	0	0	0	0	0
T600	0	0	0	0	1	0	0	0	0	0
T700	0	1	0	0	0	0	0	0	0	0
T800	1	1	0	1	1	0	1	0	1	0
T900	1	1	0	0	1	0	0	0	0	0

From Table 1-itemsets we construct Table of 3-itemset by using AND logic operation between columns of Table 1-itemsets.

For example, from column A which is (100110111) and column B which is (111101011) and column C which is (001011111), we will construct column ABC in Table of 3-itemset by applying AND logic operation on column A, column B and column C. the result will be column ABC which is (00000011) of Table 3-itemsets as in Table (2-16). After that, each column in the Table (2-16) is checked. In the case that the number of 1s is less than the minimum support threshold, then this column is neglected as in Table (2-16) and proceed to the next step [60].

Table (2-16) Table of 3-itemsets

TID	ABC	ABD	ABE
T100	0	0	1
T200	0	0	0
T300	0	0	0
T400	0	1	0
T500	0	0	0
T600	0	0	0
T700	0	0	0
T800	1	0	1
T900	1	0	0

Then each column in the Table of 2-itemsets is checked with the columns in the Table of 1-itemsets that are a subset of it. In the case of matching the columns, the correlation is extracted, as will be explained.

For example, the column (AB) in the Table of 2-itemsets is checked with columns (A) & (B) in the Table of 1-itemsets. If the column (AB) matches any of columns (A) or (B) then the association rule is extracted, otherwise, it is neglected and then proceed to the next step.

For example, the column (AE) in the Table of 2-itemsets is checked with columns (A) & (E) in the Table of 1-itemsets. the column (AE) of 2-itemsets matches column (E) of 1-itemsets expressed as follow:

The column (AE) matches the column (E) then E → A is association rule, this means that whenever appears the item E in the transaction of database, the item A surely appears in the same transaction [60].

Algorithm (2-4): ARAND Algorithm, finds association rule based on logical operation AND [60].

Input: Database of transaction (D) minimum support threshold (Th), number of column name (NC);

Output: Association rules;

Begin

Create 1-itemset-Table as Select * from D;{database transaction in binary form}

For k=2 to NC

Begin

Create k-itemset Table from combination of 1-itemset-Table ;

For i = 1 To number of columns(k-itemset-Table)

Begin

For j = 1 To number of columns ((k-1)-itemset-Table)

Begin

IF $Item_{i.}$ k-itemset-Table = $item_{j.}$ (k-1)-itemset-Table

AND [Sum($Item_{i.}$ k-itemset-columns) > = Th

AND Sum ($item_{j.}$ (k-1)-itemset- columns)] > = Th

Then

Extract Association Rule $item_{i.} \rightarrow item_{j.}$;

End if

End;

End;

Drop Table (k-1)-itemset-Table;

End;

End.

2.5.3 Association rules mining in smart environments

In the rapid development that has occurred in computer technologies and artificial intelligence, smart environments that depend on many sensors have started to appear widely, and that the mining of smart environment data is considered one of the most

important and modern types of data mining. Extracting patterns from internet of things data has emerged as important issue because of its applications in smart buildings automation [61].

One of the most important machine learning techniques is association rule mining, and association rule mining is the process of extracting important and non-trivial associative relationships between items or features in huge data within any environment.

In recent years, many researchers have studied various association rule mining Algorithms that operate on traditional transaction datasets (market-basket transactions) where unique correlations are extracted between items.

The researchers note that there has been a much increased interest in the association rules mining recently in smart environments that based sensor, in order to make important decisions in those environments that are about more comfort, energy reduction, and even this level of security. On the other hand, in other domains, such as the Internet of Things (IoT), where data is increasing daily, it needs scalable solutions in order to find smart solutions [62].

For example, in smart buildings, it is very important and useful to identify patterns or associative relationships of internet of things (IOT) devices. Where extracting these patterns and associative relationships is very useful for making strategic decisions and smart solutions from several aspects, including ease, comfort, security, reduce energy consumption and automating many devices to work together at the same time or vice versa [63].

Since the task is impossible to connect everything on earth with the Internet, here comes the role of the Internet of Things (IOT). It seems that Internet of things and sensor data in smart environments have a very big role on our lives in the future so that many things will become possible after we used to see them as impossible. With regard to the huge data that is generated from Internet of things networks and sensors

readings, surely this huge data contains useful, important and undetected hidden knowledge, so here comes the role of mining and analysis in big data in order to discover non-intuitive hidden knowledge to develop these smart systems and make them more intelligent [64].

2.5.4 Distributed association rules

Association Rule Mining (ARM) is the most popular and well studies of data mining for extracting interesting correlations among things in huge datasets. the purpose of extracting association rules is to determine important and strong correlations among various variables in huge data. Association Rule Mining have different measures of interestingness to evaluate the importance of extracted rules. Most of the classical association rule mining Algorithms focus on the sequential or centralized systems i.e. without the required communication complications [65].

Most of the existing association rules mining Algorithms work on a single computer, i.e. centralized mining, but recently the amounts of data in institutions for all disciplines have started to increase dramatically day after day. This is why it has become necessary and urgent to use several computers for the process of association rules mining, especially in organizations that have computers distributed in multiple geographic locations. For this urgent reason, Distributed Association Rule Mining (DARM) Algorithms have been designed to solve the problem of massive data growth and the existence of institutions with multiple geographic locations containing many computers in each location [65].

The purpose of distributed association rule mining Algorithms is to mine association rules on several computers located within multiple geographic locations as in Figure (2-8). Whereas, the mining process in a distributed environments require an external connection throughout the entire process. Whereas, the methods of data distribution greatly affect the efficiency of distributed association rule mining Algorithms.

There are two main classes of distributed database mining Algorithms. The first class (DARM) Algorithms sends the raw data from each site to the controller site to calculate the global association rules for all sites.

The second class of (DARM) Algorithms, which is preferred, as the mining process takes place at each site, and only the association rules are sent from each site to the controller site to calculate the global association rules for all sites as explained in Figure (2-9).

Count Distribution Algorithm (CDA), Fast Distributed Mining (FDM) Algorithm and Optimized Distributed Association Mining (ODAM) Algorithm are considered as main classification of the Classical Algorithms used in DARM [65].

Distributed association rules mining Algorithms have been developed by many researchers in the recent period, but they still suffer from some disadvantages such as time, complexity, and the loss of some important association rules [66].

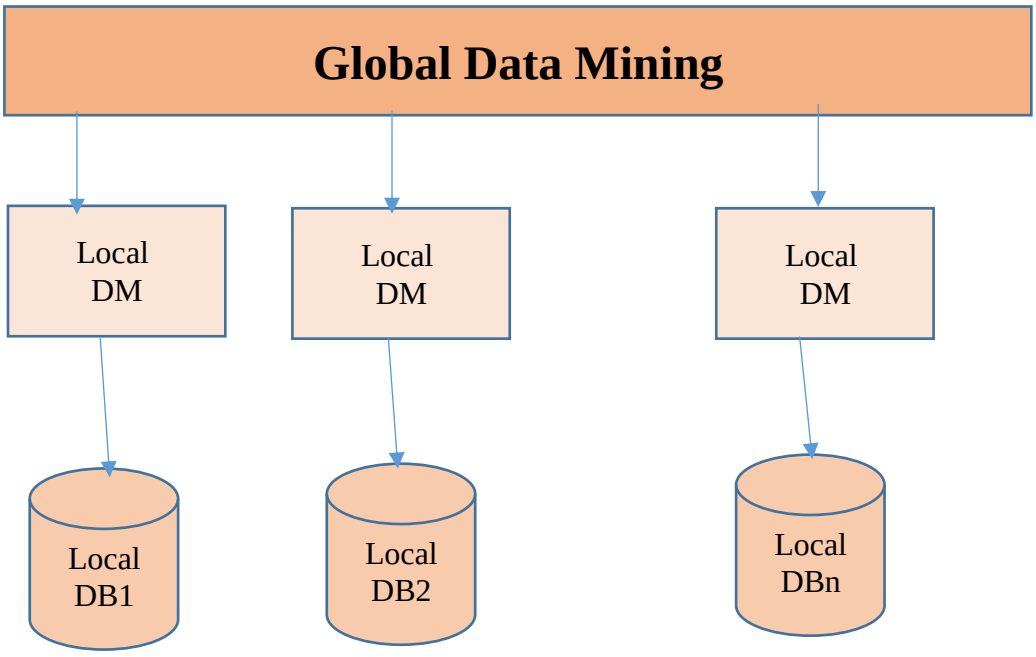


Figure (2-9) Distributed association rules [65]

2.6 Privacy preserving association rules mining over distributed environments

In various applications, digital data and information have increased dramatically in recent times, and this huge increase in the amounts of data generated represents a challenge to the efficiency of data mining [67]. Where it collects of millions of transactions of data in many disciplines such as include shopping pattern, criminal file, medical histories and acknowledgement records among others.

Distributed database, mean the data is distributed among several computers located in separate geographical locations and it have connection means among them, as explained in Figure (2-10) [68].

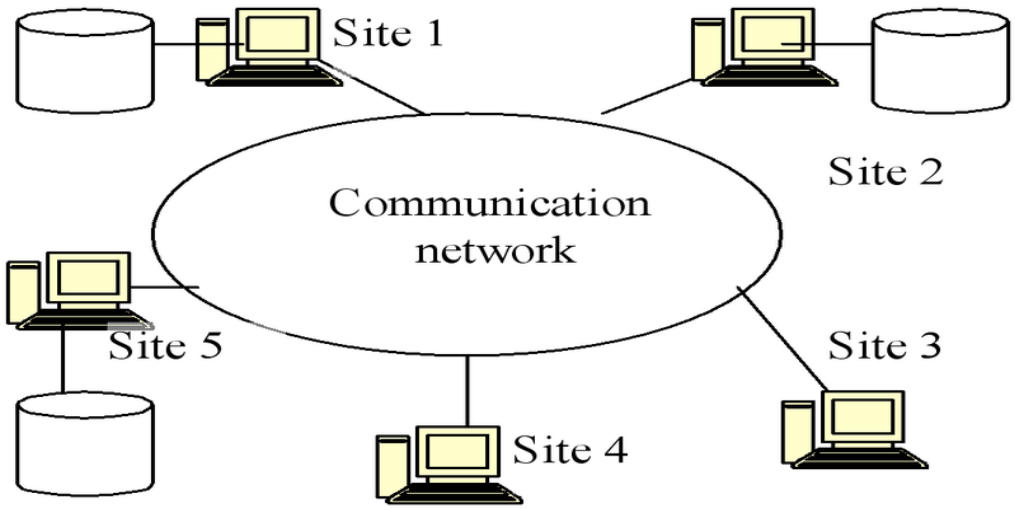


Figure (2-10) Simple distributed databases connect by communication network

In a distributed datasets setting, several data mining tasks can be done professionally and successfully. Where in last years, distributed data mining field has therefore gained great importance [69].

Data mining across different sources of data is very useful in order to extract correlations, patterns, trends, and dependencies in the datasets. The most important

type of data mining to find valuable correlations among things is association rules mining [70].

In the distributed association rules mining field, where multiple parties cooperate in order to perform the mining process on the collected data, preserving the security and privacy of the data to be mined is an important and crucial issue. As the parties do not wish to disclose their data to other parties, especially sensitive data, to preserve the privacy of individuals and institutions [71]. Privacy-preserving distributed association rule mining (PPDARM) solves this issue by mining the association rules while preserving the privacy [72]. Vertical distributed association rules, horizontal distributed association rules and hybrid distributed association rules are considered the most important types of privacy preserving of distributed association rules mining as explained in Figure (2-11).

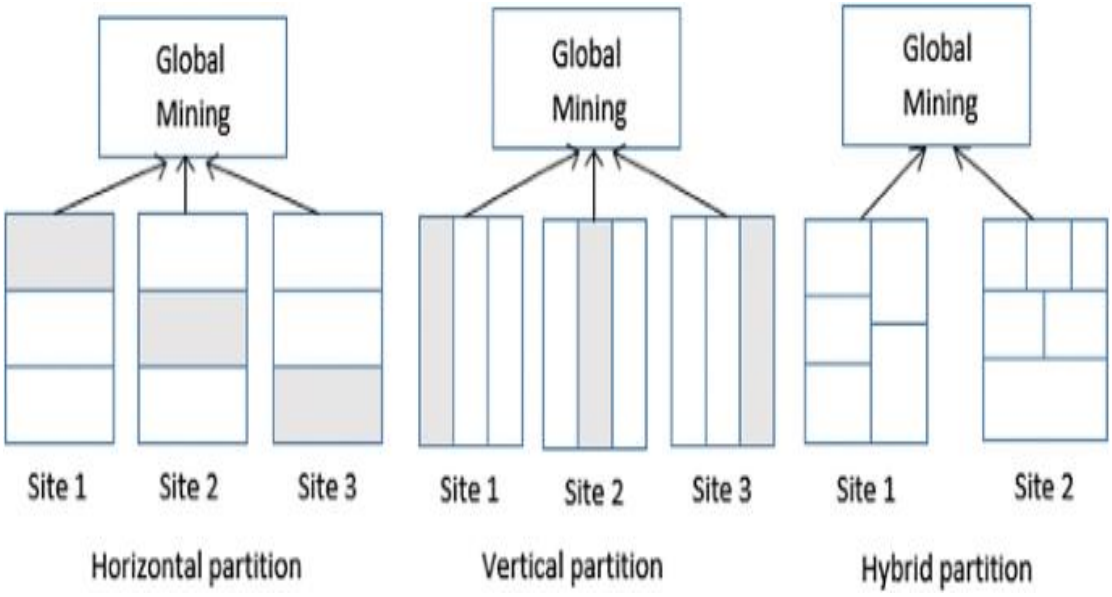


Figure (2-11) privacy preserving distributed association rule mining [46]

The majority of the privacy preserving distributed association rule mining techniques that exist suffer from many disadvantages, the most important of which are computational complexity, time, and weakness to ensure privacy for data owners [71].

Perturbation based, k- anonymity based and Cryptographic based techniques are considered the most important types of privacy preserving association rules mining techniques. As for privacy preserving association rules mining techniques in distributed environments, Cryptographic based techniques are considering the most appropriate solutions [73].

2.7 Base64 Algorithm

The Base64 Algorithm is one of the Algorithms for Encoding and Decoding an object into ASCII format, which is meant for the base number 64 or one of the methods used to encode the binary data. Base64 Commonly used in various applications such as e-mail via MME, XML data, or for URL encoding purposes. The encoding principle is to select a collection of 64 printable characters, so data can be stored and transferred across media designed to handle text data, another use of Base64 encoding is to obfuscate or randomize data. Base64 encryption schemes are usually also used when a password is needed against binary data designed to handle text-shaped data, which is intended to preserve data during transmission to a server. The characters generated by this Base64 transformation consist of A.Z, a..z and 0..9, and attached to the last two characters symbolized + and/and one character equal to (=) used for adjustment and fitting Binary data or the term is applied to as filler fitting as in Table (2-17). The character of the symbol to be generated will depend on the running Algorithm process [74]. Base64 cryptography is widely used in the internet world as a medium data format to send data, this is because the result of Base64 form is plaintext, then this information will be much easier to send, compared to the format of information in the form of binary, for the index value of the base64 Algorithm can be seen in Table 1 below:

Table (2-17) Base64 Index Value

Index	Value	Index	Value	Index	Value
0	A	**23**	X	**46**	u
1	B	**24**	Y	**47**	v
2	C	**25**	Z	**48**	w
3	D	**26**	a	**49**	x
4	E	**27**	b	**50**	y
5	F	**28**	c	**51**	z
6	G	**29**	d	**52**	0
7	H	**30**	e	**53**	1
8	I	**31**	f	**54**	2
9	J	**32**	g	**55**	3
10	K	**33**	h	**56**	4
11	L	**34**	i	**57**	5
12	M	**35**	j	**58**	6
13	N	**36**	k	**59**	7
14	O	**37**	l	**60**	8
15	P	**38**	m	**61**	9
16	Q	**39**	n	**62**	+
17	R	**40**	o	**63**	-
18	S	**41**	p		
19	T	**42**	q		
20	U	**43**	r		
21	V	**44**	s		
22	W	**45**	t		

2.8 Huffman coding

Huffman coding is very popular in data compression area. Nowadays it is used in data compression for wireless and sensor networks [75,76], data mining [77,78]. It is also found efficient for data compression in low resource systems [79]. The use of Huffman code in word-based text compression is also very common. Huffman principle produces optimal code using a Binary tree where the most frequent codewords are smaller

in length. However, Huffmanm principle does not produce a balanced tree. For this reason, it requires more memory to store longer codeword, and thus it also requires more time to decode those codewords from the memory [70].

Chapter three
Enhancement of Privacy Preserving Association Rules Based on Compression and coding Techniques Proposed system (EPPAR-Cc)

3.1 Introduction

In the contemporary world, the amount and importance of data produced by any environment in various fields is increasing. In most cases, this data needs a huge storage space for the purpose of storing and processing it in order to benefit from it, as well as the processing time may be large Relative to the huge volume of data.

Recently, data streams such as sensor data streams appeared as new systems that have its own advantages, through which it can process very large data without the need for a huge storage space and a long time for processing.

Data mining is one of the most important tools through which huge data is processed to obtain very important knowledge and is considered in most cases sensitive.

Association rules mining is one of the most important data mining techniques, through which the important correlative relationships between the items or features within the data are revealed. These correlative relationships are sensitive knowledge, so the non-leakage of this sensitive knowledge to the public or the competitors is a critical issue for data owners. Especially mining of association rules in smart environments.

In order to maintain the security and privacy of smart environments and in order to assist in making important decisions for these environments, it has become an urgent need to design a new distributed system for data mining. The proposed system obtains sensitive knowledge through the use of modern data mining techniques. it uses privacy preservation methods based on compressing and encoding to prevent the leakage of sensitive knowledge to unauthorized persons. The proposed system

relies on sensor data streams without need to store these data in order to drastically reduce storage space as well as processing time.

3.2 Design and Implementation of (EPPAR-CC)

The proposed system in terms of the structure consists of two phases. The first phase includes several distributed sites (site 1, site 2, ….., site n) executed in the same time. The second phase contains the controller site as in Figure (3-1).

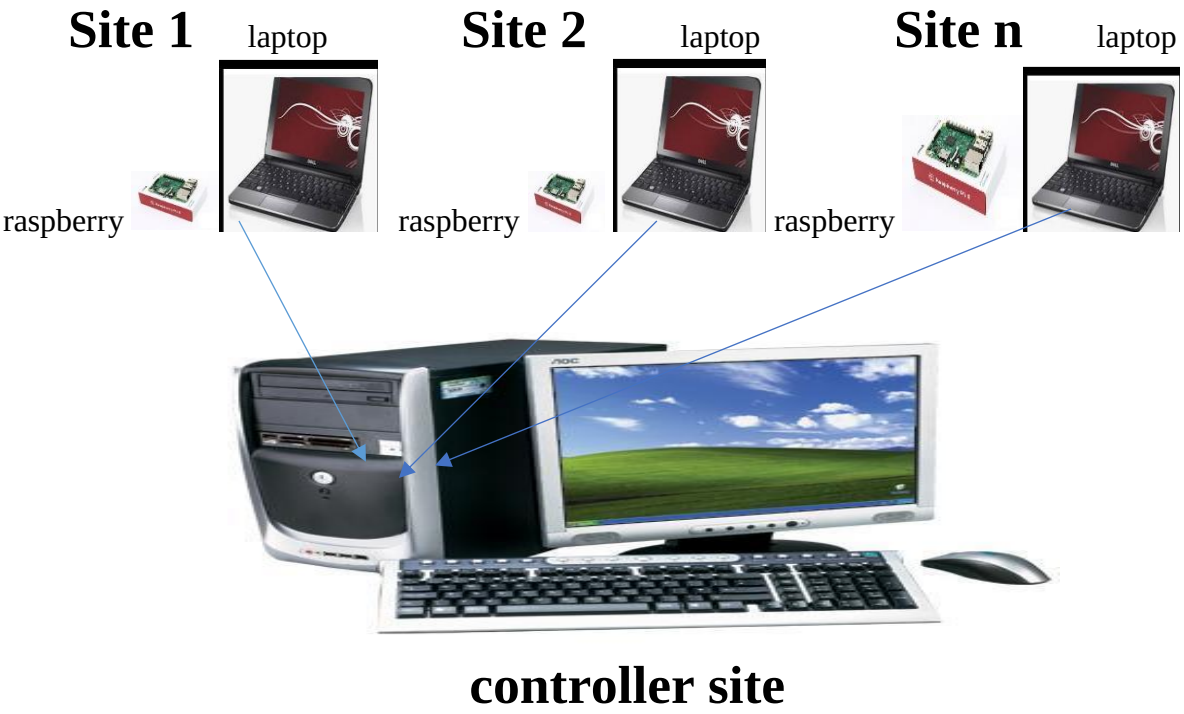


Figure (3-1) The main structure of proposed system

Phase one of the proposed system includes several Algorithms, The first Algorithm is data streams mining Algorithm that works in each site at the same time and after obtaining the results from mining process, two Algorithms are used to protect the mining results, where in each site one of the two Algorithms is used.

The first Algorithm performs data mining on sensor data streams on each site by using a proposed Algorithm of association rules mining on the streaming data through one scan for each record without need to store the streaming data. The

proposed association rule mining Algorithm works in all sites in the same time to produce many records of association rules for each site (association rules of site 1, association rules of site 2, ..., association rules of site n) as explained in Figure (3-2).

The second Algorithm in the proposed system receives the result from the first Algorithm, which includes many records of the association rules and the reduction operation is performed on them and then the hiding process also is performed. As the hiding processes used differ from one site to another. Where two different Algorithms were designed and used to hide the knowledge extracted in different sides, in order to increase the security and privacy of the extracted knowledge and not disclose it except to authorized persons because it is considered sensitive information. The output from the hiding Algorithms will be only small text is incomprehensible that represents all the associative relationships of a specific site. Hiding operations of association rules are performed in the same time on all sites and the output is sent to the controller site as explained in Figure (3-2).

The phase two of the proposed system is performed in the controller site, where the extracted knowledge from all the distributed sites is delivered, After that, the process of association rules unification for all sites is carried out to produce association rules for the whole distributed system, as each association rule within global association rules is originally present within association rules for one or more of distributed sites as explained in Figure (3-2)

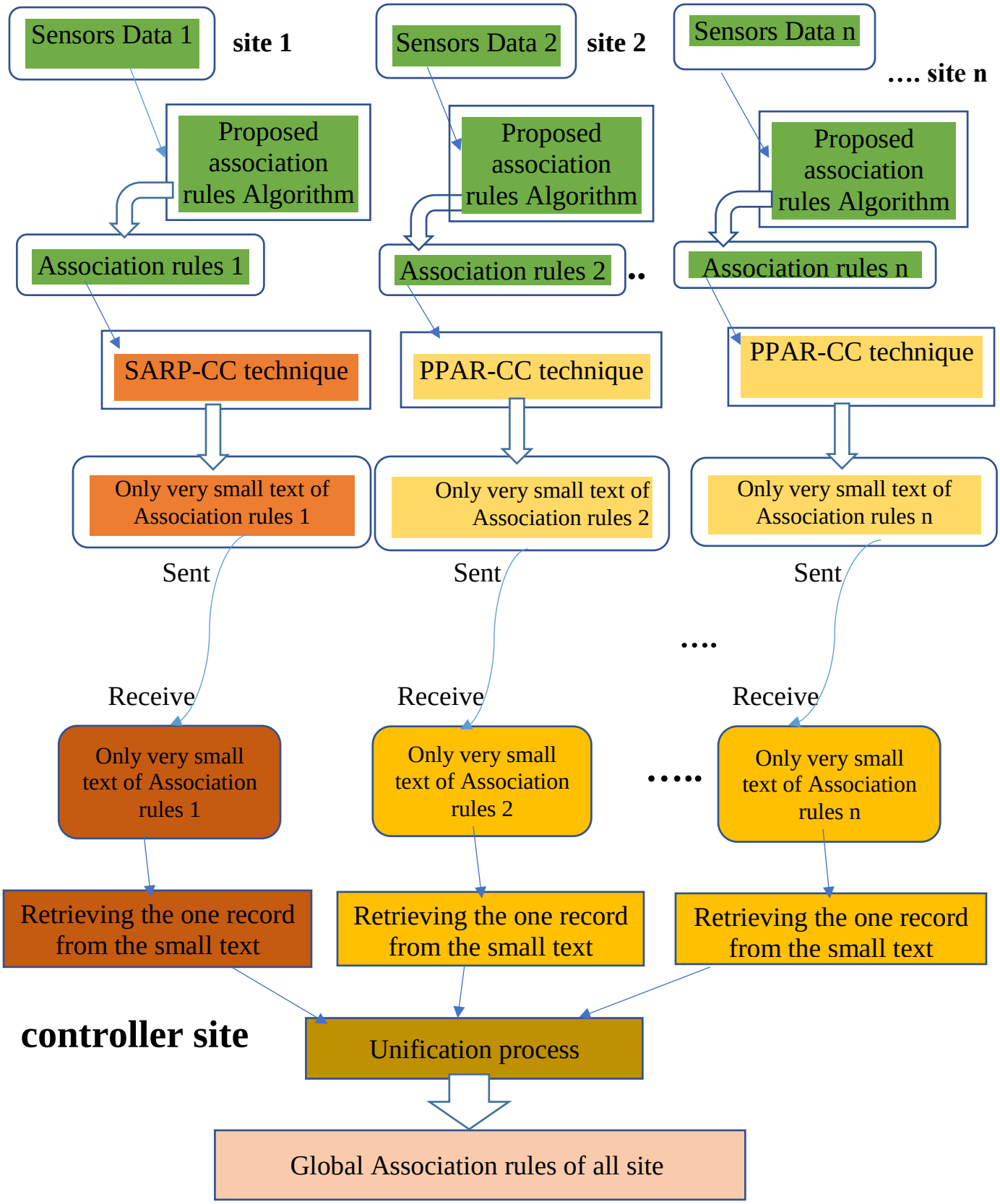


Figure (3-2) The main block diagram of proposed system

3.3 proposed system steps

The proposed system includes two phases as follows: -

1- **Phase one: Association Rules Extraction in each site in safe way.**

   A- Phase one / step one: Data streams mining (association rules mining) in each site.

   B- Phase one / step two: Compressing and Coding the knowledge (association rules).

2- **Phase two: Association Rules Extraction for all system in controller site.**

   A- Phase two / step one: Retrieving the one record from the small text.

   B- Phase two / step two: Unification of association rules of all sites.

   C- **Phase two / step three**: Extract Global Association Rules of all system.

3.3.1 Phase one: Association Rules Extraction in each site in safe way

A- Data streams mining (association rules mining) in each site

In phase one, the mining process of sensor data in each site is done in the same time through the use of the proposed association mining Algorithm called Data Stream Time Based Association Rules Mining (DST-B-ARM) that deals with time during the mining process and finding frequent itemsets and then uses the important association rules are extracted from frequent itemsets. All of this is done without the need to store the streaming data.

(DST-B-ARM) Algorithm

The main steps of (DST-B-ARM) Algorithm consists as follow:

a- find frequent itemsets

The idea behind this proposed Algorithm depends on finding frequent itemsets from sensor data streams by finding frequent 1-itemset of each sensor by one scan only and using the different structure (sensor A, sensor reading times, frequency) from each record over sensor data streams without need to store any record from streaming data, then find 2-itemset by using intersection between any two sensors such as sensors AB, (reading times of sensor A (intersection) ∩ reading times of sensor B), then to find k-itemsets by intersection process between k-1itemsets and 1-itemsets, such as sensors ABC, (reading times of sensors AB (intersection) ∩ reading times of sensor C), and so on to find all k-itemsets after pruning process for each iteration for all k-itemsets less than minimum support threshold as in Figure (3-3).

b- finding strong association rules

Extracting important association rules from frequent itemsets based on checking identical rows of frequent k-itemsets with k-1itemsets which is subset of it, then extracting association rules from identical rows.

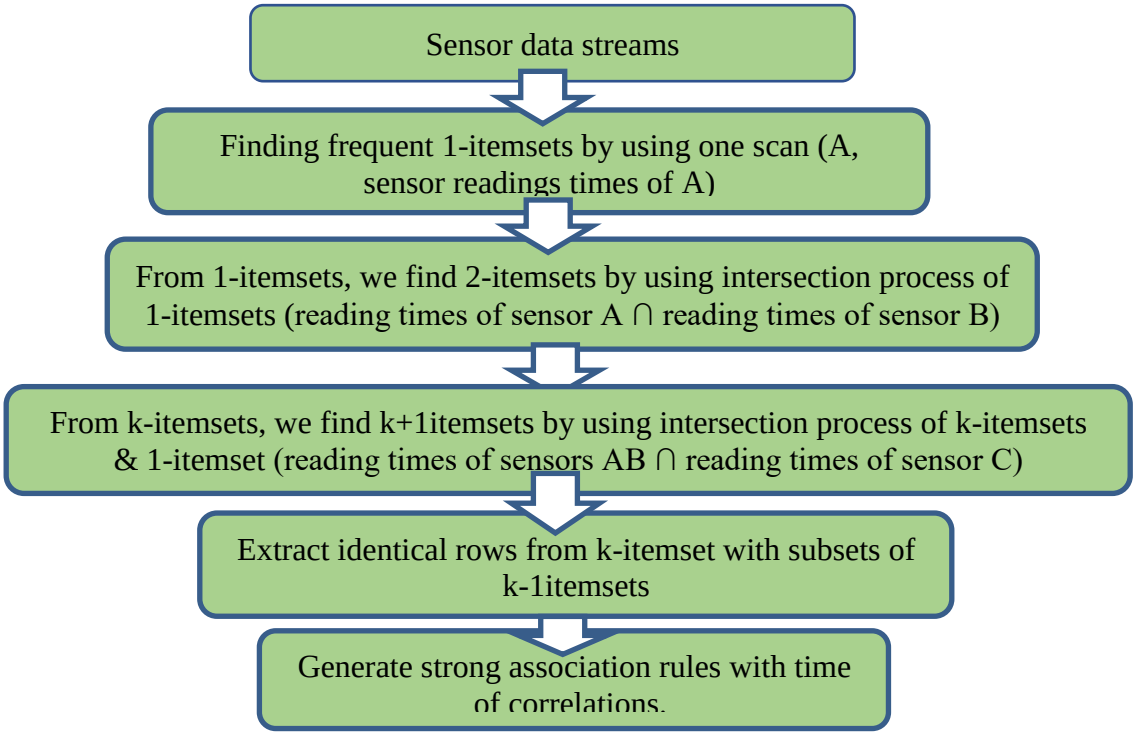


Figure (3-3) Block diagram of (DST-B-ARM) Algorithm

After applying (DST-B-ARM) Algorithm to all sites in the same time, we get many records of association rules for each site. These records of association rules are entered as input to the next Algorithms.

Algorithm (3-1): DST-B-ARM Algorithm

Input: sensor data streams with time of sensor readings, support threshold.

Output: Association rules with time periods of correlations.

Begin

A- Extracting frequent itemsets

1- Scan each record in sensor data streams once without store any record.
2- From each record extract 1-itemsets based on frequency of times of sensor readings.
3- If each 1-itemset >= support threshold then 1-itemset is frequent.
Else 1-itemset is unfrequent and deleted
4- Extract 2-itemsets from frequent 1-itemsets by using intersection process
5- If each 2-itemset >= support threshold then 2-itemset is frequent.
Else 2-itemset is unfrequent and deleted
6- If k >= 2
7- For k = 2 to no frequent itemsets of a certain size are found.
8- Extract k+1itemsets by intersection process between k-itemsets 1-itemset complement.
9- If each k+1itemset >= support threshold then k+1itemset is frequent.
10- Else k+1itemset is unfrequent and deleted
11- K=k+1
12- If no frequent itemsets of a certain size are found, then stop.

B- Extracting association Rules

13- Checking If k-itemsets = k-1itemset that is subset of it then k-itemsets & k-1itemset are identical.
14- If identical letters in k-itemsets & k-1itemset then
Put them in (association from) field
Else Put them in (association to) field

End.

B- Compressing and Coding knowledge

in this part we can use more than one Algorithm to hide the extracting knowledge (association rules). These Algorithms implemented in the same time for all sites in distributed system.

These Algorithms have the same main structure but they use different techniques in the details.

Two different Algorithms are designed to protection extracted knowledge (association rules).

The first Algorithm called Sensitive Association Rules Protection based on Compressing and Coding (SARP-CC).

The second Algorithm called Privacy Preserving Association Rules based on Compression and Coding (PPAR-CC).

(SARP-CC) Algorithm

(SARP-CC) technique used to protect sensitive knowledge (association rules) by changing the representation of knowledge from its known form to another, where it is small incomprehensible text only, so that it can be sent over distributed system.

In the beginning we perform switching the representation of association rules from their known form to another representation and placed in only one record, then reduction process is performed by several ways to small record and then finally this small record converted into incomprehensible small text to send it over the network to save privacy the data. The proposed technique (SARP-CC) includes many steps as the following:

Step 1: convert the representation of association rules from their known form ($A \rightarrow B$) to another representation and placed in only one record for each site as will explain later.

1- Putting each *k*-itemset with *k*+1itemset without any duplication in new Table and each itemset has value of two numbers.
2- Computing frequency (relations) of each *k*-itemset, with *k*-1 itemset and putting the frequency on the left side of value, except 1-itemset, it contains only the right side.
3- Computing frequency (relations) of each *k*-itemset with *k*+1-itemset and putting the frequency on the right side of value, except maximal *k*-itemset, it contains only the left side.
4- Put these association rules in one record only which represents each itemset with number of relations (L&R number) where L= represents number of relations to this itemsets with *k*-1-itemsets, R= represents number of relations to this itemsets with *k*+1-itemsets.
5- If itemsets have relations only with *k*-1itemsets, without *k*+1itemsets then putting minus before the number of relation of left side without putting right side of value.
6- If itemsets have relations only with *k*+1itemsets, without *k*-1itemsets then putting only the number of relation of right side without putting left side of value.

Step 2: Reducing fields of one record resulting from step one farther (horizontally) by using many methods serially as the following:

1. Based on prefix the subsets and their supersets (*k*-itemset with *k*+1-itemsets).
2. Based on suffix the subsets and their supersets (*k*-itemset with *k*+1-itemsets).
3. without any consider the prefix or suffix (*k*-itemset with *k*+1-itemsets).

Step 3: Locations change randomly while saving the original locations and size.

Step 4: Converting the small record by using Base64 Algorithm into very small and incomprehensible text, but rather than equal (=) put micron symbol (µ) and

double equal (==) put Beta symbol (β). Then send the small text of knowledge into controller site.

Algorithm (3-2): SARP-CC Algorithm

Input: Identical rows Table of association rules

Output: Unintelligible small text

Begin

1- Sorting all itemsets in identical rows based on k-itemsets & alphabetic

2- Put sorting itemsets in one record without any duplication.

3- **If** k-itemsets have relations with k+1itemsets & k-1itemsets then

Each field = two code value (left side value & right side value)

Left side value = number of relations satisfy that itemsets K-1itemsets

Right side value = number of relations satisfy that itemsets K+1itemsets

4- **If** k-itemsets have relations with k+1itemsets only then

Each field = one code value only

5- **If** k-itemsets have relations with k-1itemsets only then

Each field = - (one code value only)

6- Merge fields k-itemsets & k+1itemsets based on suffix

7- Merge fields k-itemsets & k+1itemsets based on prefix

8- Merge fields k-itemsets & k+1itemsets

9- Save each field location and its size

10- Change field locations randomly

11- Coding all itemsets and its orginal locations in small text based on base64

12- replace each (=) by (μ) & (==) by (β)

End.

(PPAR-CC) Algorithm

(PPAR-CC) technique is propose approach to hiding association rules after performing mining process by changing the representation of knowledge from its known form to another, where it is small text only

The proposed technique works after performing data mining process and obtaining the extracted knowledge (sensitive rules) and there is a need to send it to other sites in unsafe environments.

The proposed technique starts by converting several records into only one record that represents all the important correlations. After that we perform reducing the number of fields within that one record. The result of the is converted into incomprehensible small text that represents all knowledge for huge datasets.

The proposed technique consists four steps as the following: -

Step1: - vertical reduction.

Through converting association rules from Table contain several records to only one record as shown in applicable example.

1. Sort each itemset of left side of association rules in ascending order and without any duplication for example (A, B, C, AB, AC, BC, ABC).

2. Putting each sorted itemset of association rules in only one record without any duplication.

3. Assign a real number (contains fractional number) to all candidate itemset of association rules, for example, 1-itemset (A=0.1, B=0.2, C=0.3, D=0. 4…. etc.) and 2-itemset (AB=1.1, AC=1.2, AD=1.3, …etc.) and 3-itemsets (ABC=2.1, ABD=2.2, …etc.) and so on.

4. Representation of relationships among itemsets through assigned values as in Table (3-1).

Table (3-1) example of one record of association rules

A	AB	ABC
0.12	1.13	2.14

Where 0.12 represent assigned value of relations of itemset A, (A) represents itemset of left side of relation, (0) represents k-itemset of left side of relation, (1) represents fractional number of assigned value of left side of relation, (2) represent fractional number of assigned value of right side of relation.

5. If the itemset has more than one correlation relation, then we put comma and after the comma we adding fractional numbers of the other relations as in Figure (3-4).

A
0.12,14

Figure (3-4) Representation of two associative rules of the same itemset

Where (0) represents k-itemset of A, (12) represents fractional numbers of first relation. (14) represent fractional numbers of second relation.

Step2: - Horizontal reduction

Reducing fields of one record resulting from step one farther (horizontally) by using many methods serially as the following:

1. Based on prefix the subsets and their supersets (*k*-itemset with *k*+1-itemsets).
2. Based on suffix the subsets and their supersets (*k*-itemset with *k*+1-itemsets).
3. Without any consider the prefix or suffix (*k*-itemset with *k*+1-itemsets).

Finally, the result from one and two steps is small one record.

Step 3: Locations change randomly while saving the original locations and size.

Step4: - Converting the small record into very small and incomprehensible text by using semi-Huffman method. Then send the small text of knowledge into controller sit.

1- Calculating the frequencies of all the items within itemsets of association rules.

2- Sorting items in ascending order depending on their frequency

3- Constructing the tree based on Huffman method.

4- Extracting binary code number for each item based on Huffman tree.

5- Converting binary code into decimal number.

6- Assigning for each decimal number less than 26 corresponding capital letter (A=0, B=1, C=2……, Z=25).

7- Analyzing each number greater than 25 in the following method

   X= Y+Y+...+Z

   X= Y* y + Z

Where X represents the decimal number greater than 25, Y represents the number summed with itself one or more times to equal X, Z represents the remainder of the number, y represents the number of times Y that added with itself to be equal X.

After applying the hiding Algorithms to the extracted knowledge in the same time, the result will be only incomprehensible small text per site. This small text represents all the important correlation relationships for this site. it will be ready to be sent to the controller site and is also sent in the same time from all sites.

At the end of phase one, the result will be we have two types of incomprehensible small text, the first small text is output from the sites that used the (SARP-CC) Algorithm and the second small text is output from the sites that used the (PPAR-CC) Algorithm.

Algorithm (3-3): PPAR-CC Algorithm

Input: Association rules Table

Output: Unintelligible small text

Begin

1- Sorting all itemsets of left side of Association rules based on k-itemsets & alphabetic

2- Put sorting itemsets in one record without any duplication.

3- Assign real number for all1-itemset of association rules 0.1,0.2,….

4- Assign real number for all k-itemset of association rules r.p,r.p,…. Where K>=2, r=k-1, p=1,2,….

5- Represents the relations based on the assigned values (r.ps)
Where p = fractional number of k-itemsets or left side of rule &
s= fractional number of 1-itemsets or right side of rule

6- **If** k-itemsets have more than one relation, then
Represents relation (r.ps,py)
where y= fractional number of right side of second rule

7- Merge fields k-itemsets & k+1itemsets based on suffix

8- Merge fields k-itemsets & k+1itemsets based on prefix

9- Merge fields k-itemsets & k+1itemsets

10- Save each field location and its size

11- Change field locations randomly

11- Coding all itemsets and its orginal locations in small text based on semi huffman coding.

End.

3.3.2 phase two: Extract Global Association Rules for all system

The clear difference between phase one and phase two is that the phase two will take place in controller site only, while phase one implemented in the same time on other sites (site 1, site 2, …., site n). Where phase two begins the moment of receipt of incomprehensible small text from each site by the controller site. Where we have a number of knowledge extracted from several sites. In fact, this knowledge is extracted from two different Algorithms.

The steps of phase two consist

A- Retrieving the one record from the small text.

by using two methods

1- Retrieving one record from small text of (SARP-CC) Algorithm.
2- Retrieving one record from small text of (PPAR-CC) Algorithm.
3- Convert each one record extracted from (PPAR-CC) Algorithm into one record extracted from (PPAR-CC) Algorithm.

B- Unification one records of all site into one record which represent global association rules of all system.

C- Global Association Rules Extraction of all system.

- Retrieving the one record from the small text

1. **Retrieving one record from small text of (SARP-CC) Algorithm**

   Retrieving one record from small text of (SARP-CC) Algorithm is performed by using the same method of base 64 method after replacing (μ) by (=) & (β) by (==).

   2- Retrieving one record from small text of (PPAR-CC) Algorithm: Retrieving one record from small text of (PPAR-CC) Algorithm is performed by using sets of keys of semi-Huffman method as we will explain in applicable example.

3- Convert one record extracted from (PPAR-CC) Algorithm into one record extracted from (PPAR-CC) Algorithm

a- Scan each field in one record extracted from (PPAR-CC) Algorithm and re-represent according to real number as follow

a.bc,bd,b….

where (a) represent k-itemset of left side of rule

for example

(0) represent 1-itemset, (1) represent 2-itemset…...

(b) represent the code number of itemset of the left side of rule.

(c) represent the code number of itemset of the right side of rule.

If there is more than one relation with (b) the left side of rule, then (d) represent the right side of rule for the another relation.

b- Re-represent one record according into steps of (PPAR-CC) Algorithm as we will explain in applicable example.

At the end of this step we get the same type for one record extracted from (SARP-CC) Algorithm.

B- Unification one records of all sites

In this step, the one record of association rules for each site is gathered in one record only that represents the knowledge (association rules) of all system sites.

Algorithm of Unification of association rules of all sites in controller site

Algorithm (3-4): - U-AR Algorithm

Input: one record of AR of site 1, one record of AR of site 2….. one record of AR of site n,

Output: AR of all site in one record

Begin

1- read one record of association rules 1

2- **For** i = 2 to n

4- read one record of association rules

5- **for** j = 1 to r

6- **if** itemset found in one record of association rules 1 then
add left value into left value of association rules 1 &
add right value into right value of association rules 1

7- else add it with it values into one record of association rules 1

8- End for j

9- AR of all site in one record

10- End for i

End.

C- Global Association Rules Extraction of all system

In this step, global association rules are extracted from the one record that represents association rules for all sites.

Algorithm (3-5): Global Association Rules Extraction (GARE) Algorithm
Input: association rules of all sites in one record
Output: Global Association Rules
Begin
1.Read one record of association rules of all sites
2.K=2
3.Read all k-itemsets
4.Read all k-1itemsets
5.Fetch first field of k-itemset and looking for k-1itemset which is subset of it.
If k-itemset have two number of value & k-1itemset have one number of value, then
If left number of value of k-itemset = the value of k-1itemset, then
left number of value of k-itemset & the value of k-1itemset = 0
If left number of value of k-itemset > the value of k-1itemset, then
left number of value of k-itemset = the difference.
If left number of value of k-itemset < the value of k-1itemset, then
the value of k-1itemset = the difference.
End if.
If k-itemset have two number of value & k-1itemset have two number of value, then
If left number of value of k-itemset = right number of value of k-1itemset, then
left number of value of k-itemset & right number of value of k-1itemset = 0
If left number of value of k-itemset > right number of value of k-1itemset, then
left number of value of k-itemset = the difference.
If left number of value of k-itemset < the value of k-1itemset, then
right number of value of k-1itemset = the difference.
End if.
If k-itemset have minus value in & k-1itemset have one number of value, then
If value of k-itemset = value of k-1itemset, then
value of k-itemset = value of k-1itemset = 0
If value of k-itemset < value of k-1itemset, then
value of k-1itemset = the difference.
If value of k-itemset > value of k-1itemset, then
value of k-itemset = the difference.
End if.
If k-itemset have minus value & k-1itemset have two number of value, then

If value of k-itemset = right side of value of k-1itemset, then
value of k-itemset = right side of value of k-1itemset = 0
If value of k-itemset < right side of value of k-1itemset, then
right side of value of k-1itemset = the difference.
If value of k-itemset > right side of value of k-1itemset, then
value of k-itemset = the difference.
End if

6. Repeat step 5 for all k-itemset
7. Create new record with remaining values.
8. K= K+1
9. Repeat 3,4,5,6,7 steps for each k-itemset

End.

Chapter four

Discussion and Experiment Results

4.1 Introduction

This chapter presents the description of the implementation of Algorithms of the distributed proposed system used to test the proposed system. This chapter consists of six sections, a brief introduction in the first section, the second section contains Datasets Layout used for implementation, the third section presents Applicable example. Section four presents The Implementation Environment. Section five shows experiments for the proposed system (Implementation in each site, and Implementation in controller site for all sites). Section six presents Discussion of Accuracy of Results. Section seven presents the Comparisons.

4.2 Datasets Layout

There are two types of possible layouts of the data set for data mining, the horizontal and vertical layouts.

The datasets used in the implementation and comparison is based on the reference [81] of a doctoral dissertation, which represents sensors readings with the time of sensors readings of 5 sensors (motion sensor (A), door opening and closing sensor (B), lighting sensor (C), temperature sensor (D), and humidity sensor (E)).

These data were used only to properly check the functioning of the proposed system, and that the proposed system could work in any smart environment that relies on sensors, or rather any environment that contains data streams with times of data streams.

4.3 Applicable example

Assume we have Raspberry have five sensors (motion sensor (A), door opening and closing sensor (B), lighting sensor (C), temperature sensor (D), and humidity sensor

(E)) and we have the following readings for the data flowing from the sensors with reading times of sensors as in Table (4-1)

Table (4-1) Sensor streaming data

tid	Sensors reading	time
1.	1,0,0,24,37	9:01
2.	0,0,1,24,37	9:02
3.	0,0,0,22,38	9:03
4.	1,0,0,21,33	9:04
5.	0,0,1,24,37	9:05
6.	1,0,0,24,37	9:06
7.	1,0,0,24,37	9:07
8.	0,0,0,23,38	9:08
9.	1,0,0,21,33	9:09
10.	0,0,1,24,37	9:10
11.	0,0,0,22,39	9:11
12.	0,0,0,24,40	9:12
13.	1,0,0,24,37	9:13
14.	1,1,0,21,33	9:14
15.	1,0,1,24,37	9:15
16.	0,0,0,22,36	9:16
17.	0,0,1,24,37	9:17
18.	1,0,0,22,39	9:18
19.	1,0,0,24,37	9:19
20.	1,0,0,25,36	9:20

4.3.1 Applying proposed system (EPPAR-CC)

Phase one: Association Rules Extraction in each site in safe way

1- Data streams mining (association rules mining) in each site.

- **(DST-B-ARM) Algorithm**

In this step, the proposed Algorithm (DST-B-ARM) is applied

Step one: Extracting frequent itemset

1- Finding frequent 1-itemset as in Table (4-2).

Table (4-2) frequent 1-itemsets

item	Times	F
A	9:01, 9:04, 9:06-9:07, 9:09, 9:13-9:15, 9:18-9:20	11
D	9:01-9:02, 9:05-9:07, 9:10, 9:12-9:13, 9:15, 9:17, 9:19-:9:20	12
E	9:01-9:03, 9:05-9:08, 9:10-9:13, 9:15, 9:17-9:19	17
F	9:02-9:03, 9:05, 9:08, 9:10-9:12, 9:16-9:17	9
G	9:01-9:13, 9:15-9:20	19
H	9:01, 9:03-9:04, 9:06-9:09, 9:11-9:14, 9:16, 9:18-9:20	15
I	9:03-9:04, 9:08-9:09, 9:11, 9:14, 9:16, 9:18	8

Table 1-itemsets is to be filled in through one scan on each record of sensor reading Table , so if sensor A gives reading 1 it adds the times to a record A in Table 1-itemsets and at the same time it adds the times when the sensor A reading is equal to 0 to the record F. This means that the times in the record F are the times Not found in record A (complementary).

Thus, the remaining records are filled in the same way, such as the record B, and the one that completes it G, the record C, and the one that completes it H, As for the case of the sensor D (temperature sensor), the value is an integer and not 0, 1 in this case we divide the numbers into the two periods upper and lower such as 23 and above are considered 1 and put the times in record D of Table 1-itemsets, while 22 and below are considered 0 and we put the times in the record I, and so on with the sensor E (humidity sensor), where times are placed in record E if the readings are 36 and above, and times are placed in record J if the readings are 35 or below. Where this Table is the frequent 1-itemsets after pruning or deleting any record less than the threshold that is 6 such as the records B, C.

2- Constructing 2-itemsets by using the intersection operation between 1-itemsets as in Table (4-3).

Table (4-3) frequent 2-itemsets

itemset	Times	F
AE	9:01, 9:06-9:07, 9:13, 9:15, 9:18-9:20	8
AG	9:01, 9:04, 9:06-9:07, 9:09, 9:13, 9:15, 9:18-9:20	10
AH	9:01, 9:04, 9:06-9:07, 9:09, 9:13-9:14, 9:18-9:20	10
DE	9:01-9:02, 9:05-9:07, 9:10, 9:12-9:13, 9:15, 9:17, 9:19-9:20	12
DG	9:01-9:02, 9:05-9:07, 9:10, 9:12-9:13, 9:15, 9:17, 9:19-9:20	12
EF	9:02-9:03, 9:05, 9:08, 9:10-9:12, 9:16-9:17	9
EG	9:01-9:03, 9:05-9:08, 9:10-9:13, 9:15-9:20	17
EH	9:01, 9:03, 9:06-9:08, 9:11-9:13, 9:16, 9:18-9:20	12
FG	9:02-9:03, 9:05, 9:08, 9:10-9:12, 9:16-9:17	9
GH	9:01, 9:03-9:04, 9:06-9:09, 9:11-9:13, 9:16, 9:18-9:20	14
HI	9:03-9:04, 9:08-9:09, 9:11, 9:14, 9:16, 9:18	8

3- Constructing 3-itemsets by using intersection process as n Table (4-4).

Table (4-4) frequent 3-itemsets

itemset	Times	F
AEG	9:01, 9:06-9:07, 9:13, 9:15, 9:18-9:20	8
AGH	9:01, 9:04, 9:06-9:07, 9:09, 9:13, 9:18-9:20	9
DEG	9:01-9:02, 9:05-9:07, 9:10, 9:12-9:13, 9:15, 9:17, 9:19	12
EFG	9:02-9:03, 9:05, 9:08, 9:10-9:12, 9:16-9:17	9
EGH	9:01, 9:03, 9:06-9:08, 9:11-9:13, 9:16, 9:18-9:20	12

Step two: Finding strong association rules

From Tables of frequent itemsets we can extract identical rows and then association rules Accompanied with the association time, which will be the first time and last time within frequent itemset as explained in Tables (4-5), (4-6).

To reduce storage space, we use time code rather than the Real or explicit time.

Table (4-5) identical rows

ID	k-itemsets	k-1itemsets	Ass. time
1	DE	D	9:01-9:20
2	DG	D	9:01-9:20
3	FG	F	9:02-9:17
4	EG	E	9:01-9:20
5	EF	F	9:02-9:17
6	AEG	AE	9:01-9:20
7	DEG	DE	9:01-9:20
8	DEG	DG	9:01-9:20
9	DEH	DH	9:01-9:20
10	DGH	DH	9:01-9:20
11	EGH	EH	9:01-9:20
12	EFG	EF	9:02-9:17
13	EFG	FG	9:02-9:17
14	DEGH	DEH	9:01-9:20
15	DEGH	DGH	9:01-9:20

Table (4-6) association rules

Ass.from	Ass.to	Time code
D	E	1
D	G	1
F	G	2
E	G	1
F	E	2
AE	G	1
DE	G	1
DG	E	1
DH	E	1
DH	G	1
EH	G	1
EF	G	2
FG	E	2
DEH	G	1
DGH	E	1

2- Compressing and Coding knowledge (association rules).

In this part we use two Algorithms (SARP-CC Algorithm and PPAR-CC Algorithm) in different site. One of these two Algorithms in each site.

- SARP-CC Algorithm

The inputs of this Algorithm are the results of (DST-B-ARM) Algorithm and the output only incomprehensible small text as in the following steps.

Step 1: convert the representation of association rules from their known form (*A* →*B*) to another representation and placed in only one record for each site as in Table (4-7).

Table (4-7) one record of association rules

D	E	F	A	D	D	D	E	E	E	F	AE	DE	DE	DG	EF	EG	DEG
2	1	2	1	11	11	2	1	-1	1	11	-1	-2	11	11	-2	-1	-2

Step 2: compress the results from step one farther as in Table (4-8).

Table (4-8) one record of association rules after fields reduction

DH	F2G		A2E2G		D2G2H		E2G2H		DE2H2		E3F2G			D4E3G2H			
2	2	11	1	-1	11	11	-1	-1	11	1	1	11	-2	2	11	-2	-2

Step 3: Locations change randomly while saving the original locations and size

Step 4: hiding the small record by using Base64 Algorithm and convert one record into incomprehensible small text, but rather than (=) put (*μ)* and (==) put (*β)*. Then send the the smll text into controller site.

“RTNGMkcμREgμRDJHMkgμREUySDIμRTJHMkgμRjJHRDRFM0cySAβQTJFMkcμ11|11-1|-111|11|11|-22|11|-2|-21|-12|112:OCw4LDgsOCwxMSw4LDQsNDoxMCw0LDUsNCw1LDQsMSw3OjAsMyw1LDIsNCw2LDcsMTowLDMsNSwyLDQs”

- The PPAR-CC Algorithm

The PPAR-CC Algorithm steps consist

Step 1: - Converting association rules from Table contain several records to only one record as follow.

1- Sort each itemset of left side of association rules in ascending order and without any duplication for example (D, E, F, I, AE, DE, DG, EF, EH, FG).

2- Putting each sorted itemset of association rules of left side in only one record without any duplication as in Table (4-9).

3- Assign a real number (contains fractional number) to each candidate itemset of all association rules, for example, 1-itemset (D=0.1, E=0.2, F=0.3, G=0.4,), 2-itemset (AE=1.1, DE=1.2, DG=1.3, DH=1.4, EH=1.5, EF=1.6, FG=1.7), 3-itemset (DEH=2.1, DGH=2.2).

4- Representation of relationships among itemsets through assigned values as in Table (4-9).

Where 0.12 represents assigned value of relations of itemset D, (D) represents itemset of lift side of relation, the first (0) represents k-itemset of lift side of relation, (1) represents fractional number of assigned value of lift side of relation, (2) represent fractional number of assigned value of right side of relation, (14) represents the second relation with itemset D.

Table (4-9) one record of association rules

D	E	F	AE	DE	DG	DH	EF	EH	FG	DEH	DGH
0.12,14	0.24	0.32,34	1.14	1.24	1.34	1.44,42	1.54	1.64	1.72	2.12	2.22

Step2: - Horizontal reduction of fields as in Table (4-10).

Table (4-10) one record of association rules after fields reduction

AE	DH	EH	E2F		F2G		D2G2H		D3E2H		
1.14	1.44,42	1.64	0.24	1.54	0.32,34	1.72	1.34	2.22	0.12,14	1.24	2.12

Finally, the result from one and two steps is small one record.

Step 3: Locations change randomly while saving the original locations and size

Step4: - converting one record of association rules into incomprehensible small text by using semi-Huffman method.

1- Calculating the frequencies of all the items within itemsets of association rules.

(A=1, D=3, E=5, F=3, G=2, H=1, I=1)

2- Sorting items in ascending order depending on their frequency

(A=1, H=1, I=1, G=2, D=3, F=3, E=5)

3- Constructing the tree based on Huffman method as in Figure (4-1)

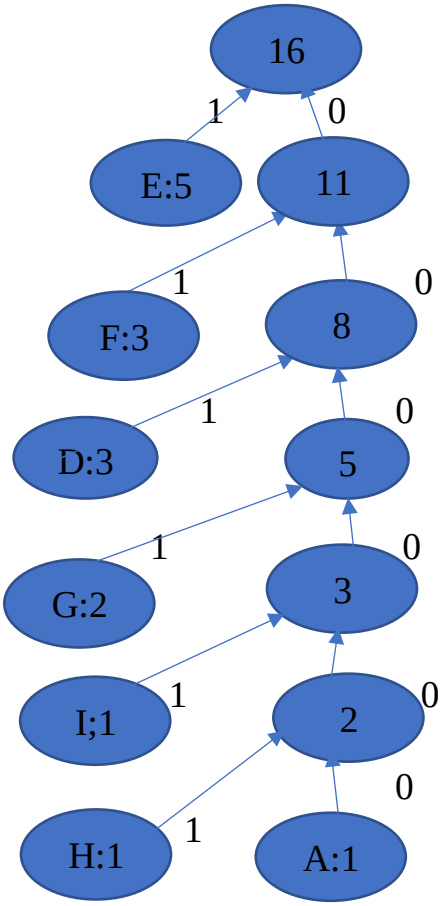


Figure (4-1) tree of coding items

4- Extracting binary code number for each item based on Huffman tree. (E=0, F=10, D=110, G=1110, I=11110, H=111110, A=111111.

5- Converting binary code into decimal number.

(E=0, F=2, D=6, G=14, I=30, H= 62, A=63).

6- Assigning for each decimal number less than 26 corresponding capital letter (A=0, B=1, C=2……, Z=25).

E=A, F=C, D=G, G=O.

7- Analyzing each number greater than 25 in the following method

8- From step 7 Assigning for each code corresponding capital letter and small letter as the following (A=0, B=1... Z=25) and the frequency by small letter. (a=1, b=2……z=26). I=15*2=Pb, H=20*3+2= Uc&B, A=20*3+3= Uc&C

"OcPCGCcOAcPcGPbCAdCcGAG0.32,34|1.721.640.24|1.541.34|2.221.140.12,14|1.24|2.121.44,42:MywyLDMsNSwzLDUsMjoxMiw0LDksOSw0LDE3LDc6Miw2LDEsMyw0LDAsNToyLDYsMSwzLDQsMCw1OjIzOjYy"

Phase two: Extract Association Rules for all system

In the beginning of phase two the controller site receives small text from different sites.

The steps of extract association rules of all sides consist: -

1- Retrieving one record from small text

- Retrieving one record from small text of (SARP-CC) Algorithm

Retrieving one record from small text of (SARP-CC) Algorithm is performed by using the same method of base 64 method after replacing (μ) by (=) & (β) by (==) as in Figure (4-2).

RTNGMkcμREgμRDJHMkgμREUySDIμRTJHMkgμRjJHRDRFM0cySAβQTJFMkcμ11|11-1 |-111|11|11|-22|11|-2|-21|-12|112:OCw4LDgsOCwxMSw4LDQsNDoxMCw0LDUsNCw1LDQsMSw3OjAsMyw1LDIsNCw2LDcsMTowLDMsNSwyLDQs

DH	F2G		A2E2G		D2G2H		E2G2H		DE2H2		E3F2G			D4E3G2H			
2	2	11	1	-1	11	11	-1	-1	11	1	1	11	-2	2	11	-2	-2

D	E	F	AE	DE	DG	DH	EF	EG	EH	FG	AEG	DEG	DEH	DGH	EFG	EGH	DEGH
2	1	2	1	11	11	2	11	-1	1	11	-1	-2	11	11	-2	-1	-2

Figure (4-2) Retrieving one record from small text of (SARP-CC) Algorithm

- Retrieving one record from small text of (PPAR-CC) Algorithm

Retrieving one record from small text of (PPAR-CC) Algorithm is performed by using sets of keys of semi-Huffman method as in Figure (4-3).

(E=A, F=C, D=G, G=O, I=Pb, H=Uc&B, A=Uc&C).

OcPCGCcOAcPcGPbCAdCcGAG0.32,34|1.721.640.24|1.541.34|2.221.140.12,14|1.24|2.121.44,42:MywyLDMsNSwzLDUsMjoxMiw0LDksOSw0LDE3LDc6Miw2LDEsMyw0LDAsNToyLDYsMSwzLDQsMCw1OjIzOjYy

AE	DH	EH	E2F		F2G		D2G2H		D3E2H		
1.14	1.44,42	1.64	0.24	1.54	0.32,34	1.72	1.34	2.22	0.12,14	1.24	2.12

D	E	F	AE	DE	DG	DH	EF	EH	FG	DEH	DGH
0.12,14	0.24	0.32,34	1.14	1.24	1.34	1.44,42	1.54	1.64	1.72	2.12	2.22

Figure (4-3) Retrieving one record from small text of (PPAR-CC) Algorithm

- Convert one record extracted from (PPAR-CC) Algorithm into one record extracted from (SARP-CC) Algorithm

To convert one record extracted from (PPAR-CC) Algorithm into one record extracted from (SARP-CC) Algorithm.

Now we have one record of association rules as in Figure (4-4).

D	E	F	AE	DE	DG	DH	EF	EH	FG	DEH	DGH
0.12,14	0.24	0.32,34	1.14	1.24	1.34	1.44,42	1.54	1.64	1.72	2.12	2.22

D	E	F	AE	DE	DG	DH	EF	EG	EH	FG	AEG	DEG	DEH	DGH	EFG	EGH	DEGH
2	1	2	1	11	11	2	11	-1	1	11	-1	-2	11	11	-2	-1	-2

Figure (4-4) Convert one record of (PPAR-CC) into one record of (SARP-CC)

2- Unification one records of all sites

In this step we use Algorithm of Unification of association rules to collect all one record of two sites in one record which represent association rules of all system in compressed form.

3- Extract Association Rules of all system

Finally, we can extract global association rules of all sites from unification of one record of association rules by using Algorithm of global association rules extraction

1- We begin with k-itemset (2-itemset) and k-1itemset (1-itemset) to extract association rules as in Table (4-11).

Table (4-11) Extracting identical rows of 2-itemset &1-itemset

DE	DG	EF	EG	FG
-1	-1	-1	-1	-1

D	E	F
2	1	2

2itemsets	1itemset
DE	D
DG	D
EF	F
EG	E
FG	F

2- Fetch 3-itemset with 2-itemset to extract association rules as in Table (4-12).

Table (4-12) Extracting identical rows of 3-itemset &2-itemset

AEG	DEG	DEH	DGH	EFG	EGH
-1	-2	-1	-1	-2	-1

AE	DE	DG	DH	EF	EH	FG
1	1	1	2	1	1	1

3itemsets	2itemset
AEG	AE
DEG	DE
DEG	DG
DEH	DH
DGH	DH
EFG	EF
EFG	FG
EGH	EH

3- Fetch 4-itemset with 3-itemset to extract association rules as in Table (4-13).

Table (4-13) Extracting identical rows of 4-itemset &3-itemset

DEGH
-2

DEH	DGH
1	1

3itemsets	2itemset
DEGH	DEH
DEGH	DGH

Finally, we have association rules of all system without any false negatives, where each association rule in global association rules is originally exist within one or more association rules of sites as in Table (4-14).

Table (4-14) association rules of all system

NO	K-itemsets	k-1itemsets		Ass.from	Ass.to
1	DE	D		D	E
2	DG	D		D	G
3	EF	F		F	E
4	EG	E		E	G
5	FG	F		F	G
6	AEG	AE		AE	G
7	DEG	DE		DE	G
8	DEG	DG		DG	E
9	DEH	DH		DH	E
10	DGH	DH		DH	G
11	EFG	EF		EF	G
12	EFG	FG		FG	E
13	EGH	EH		EH	G
14	DEGH	DEH		DEH	G
15	DEGH	DGH		DGH	E

4.4 The Implementation Environment

The standard implementation environment depends on at least three computers and two raspberries, as each computer and its raspberries are considered a site and the third computer is considered the controller site to calculate the association rules for the system as a whole.

The feature of each computer are: (Intel(R) Core(TM) i3-2350M CPU @ 2.30GHz 2.30 GHz, RAM 4.00 GB, 64-bit operating system, x64-based processor).

The proposed system is implemented by using C# programming language.

4.5 Experiments

Consider the distributed system which consists of two branches at different sites, S1and S2. There are 43,200 records of 12 hours in each site, with 5 sensors & 10

features in each site Figure (4-5). There are experiments for implementation at every site, as well as implementation in control site for the system as a whole.

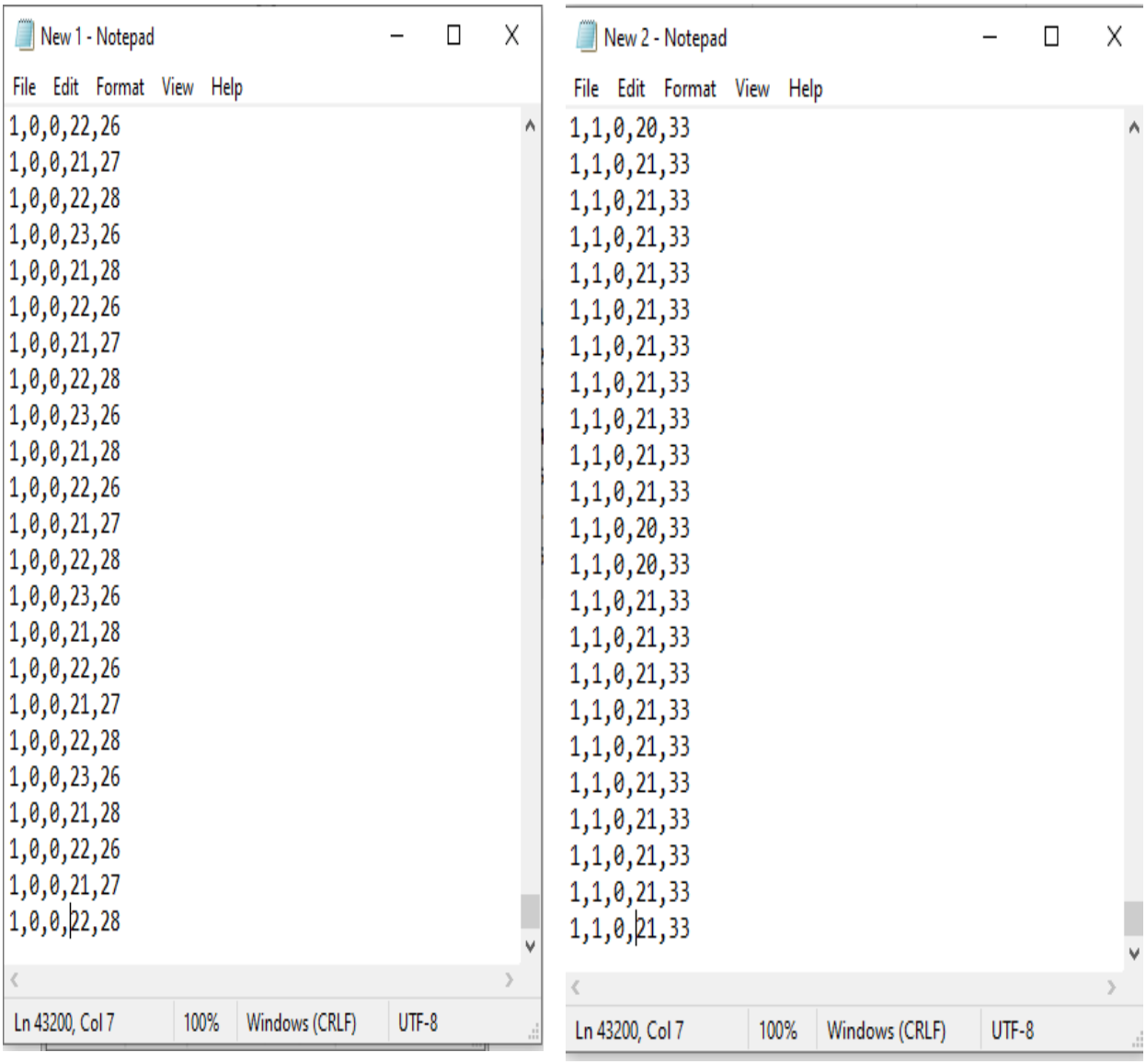


Figure (4-5) Dataset used in Implementation of site 1 & site 2

4.5.1 Implementation of proposed system (EPPAR-CC)

- Local implantation

A- Local implantation of site 1

1- Implantation of DST-B- ARM Algorithm.

2- Implantation of (SARP-CC) Algorithm.

Table (4-15) and Figure (4-6) represent the implementation of proposed system in site 1, as explained in proposed system steps (phase one) in chapter three.

B- Local implantation of site 2

1- Implantation of DST-B- ARM Algorithm.

2- Implantation of (PPAR-CC) Algorithm.

Table (4-16), Figure (4-7) represent the implementation of proposed system in site 2, as explained in proposed system steps (phase one) in chapter three, as a result of the implementation in site 1 & site 2 (small text) is now sent to controller site from each site (1,2) which represents association rules for site 1 & association rules for site 2.

- Implantation of DST-B- ARM Algorithm of site 1

Table (4-15) Association rules with association times of site 1

No	K-itemset	k-1itemset	Association time	Time code	Ass. from	Ass. to
1	FH	F	01:06:17-03:50:31, 06:28:47-09:3...	1	F	H
2	EHI	EH	01:06:40-01:16:10, 06:28:47-13:0...	2	EH	I
3	FHI	FI	01:06:26-03:50:31, 06:28:47-09:3...	3	FI	H
4	HIJ	IJ	01:06:26-06:28:46, 09:34:14-11:1...	4	IJ	H
5	AHIJ	AIJ	01:16:11-02:24:49, 03:50:32-06:2...	5	AIJ	H
6	GHIJ	GIJ	01:06:26-06:28:46, 09:34:14-11:1...	6	GIJ	H

- ***Site1 DST-B- ARM Algorithm implementation***
  - Number of association rules with association times is (6).
  - Execution time is (1.30) minutes.
  - Storage Space is (50) KB.

- Implantation of (SARP-CC) Algorithm

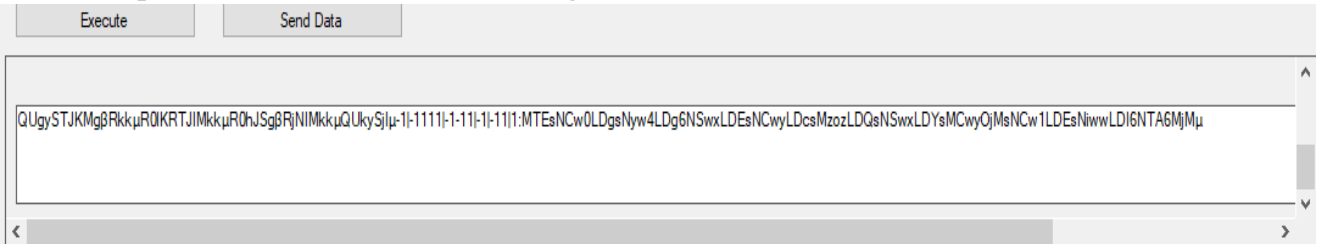
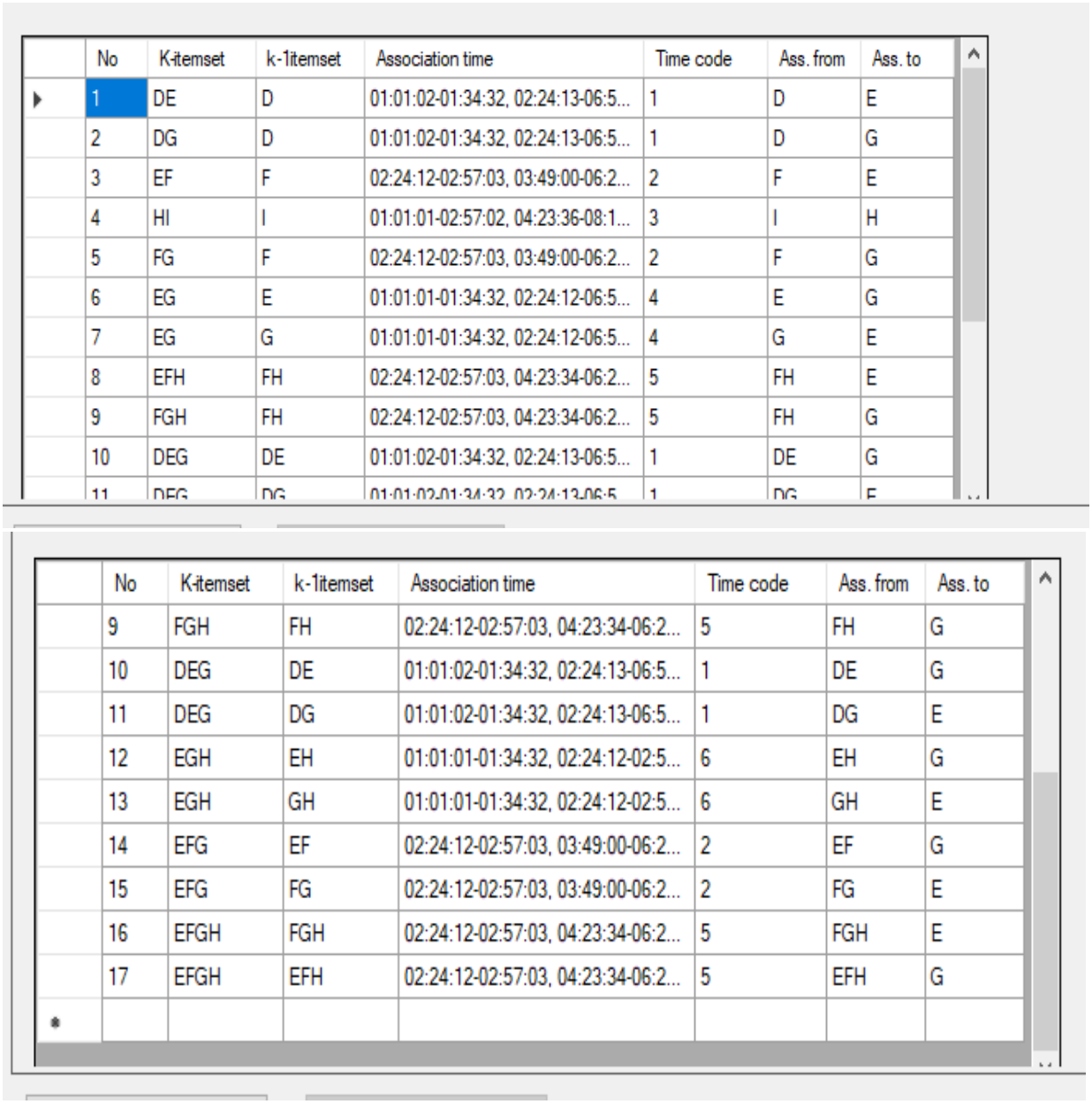


Figure (4-6) Change locations randomly then convert one record into incomprehensible small text of site 1

- ***Site1 (SARP-CC) Algorithm implementation***
  - The output is very small and unintelligible text (only one line).
  - Execution time is less than (1) second.
  - Storage Space is (12) KB.

- Implantation of DST-B- ARM Algorithm of site 2

Table (4-16) Association rules with association times of site 2



No	K-itemset	k-1itemset	Association time	Time code	Ass. from	Ass. to
1	DE	D	01:01:02-01:34:32, 02:24:13-06:5...	1	D	E
2	DG	D	01:01:02-01:34:32, 02:24:13-06:5...	1	D	G
3	EF	F	02:24:12-02:57:03, 03:49:00-06:2...	2	F	E
4	HI	I	01:01:01-02:57:02, 04:23:36-08:1...	3	I	H
5	FG	F	02:24:12-02:57:03, 03:49:00-06:2...	2	F	G
6	EG	E	01:01:01-01:34:32, 02:24:12-06:5...	4	E	G
7	EG	G	01:01:01-01:34:32, 02:24:12-06:5...	4	G	E
8	EFH	FH	02:24:12-02:57:03, 04:23:34-06:2...	5	FH	E
9	FGH	FH	02:24:12-02:57:03, 04:23:34-06:2...	5	FH	G
10	DEG	DE	01:01:02-01:34:32, 02:24:13-06:5...	1	DE	G
11	DEG	DG	01:01:02-01:34:32, 02:24:13-06:5	1	DG	E

No	K-itemset	k-1itemset	Association time	Time code	Ass. from	Ass. to
9	FGH	FH	02:24:12-02:57:03, 04:23:34-06:2...	5	FH	G
10	DEG	DE	01:01:02-01:34:32, 02:24:13-06:5...	1	DE	G
11	DEG	DG	01:01:02-01:34:32, 02:24:13-06:5...	1	DG	E
12	EGH	EH	01:01:01-01:34:32, 02:24:12-02:5...	6	EH	G
13	EGH	GH	01:01:01-01:34:32, 02:24:12-02:5...	6	GH	E
14	EFG	EF	02:24:12-02:57:03, 03:49:00-06:2...	2	EF	G
15	EFG	FG	02:24:12-02:57:03, 03:49:00-06:2...	2	FG	E
16	EFGH	FGH	02:24:12-02:57:03, 04:23:34-06:2...	5	FGH	E
17	EFGH	EFH	02:24:12-02:57:03, 04:23:34-06:2...	5	EFH	G

- ***Site2 DST-B-ARM Algorithm implementation***
  - Number of association rules with association times is (17).
  - Execution time is (1.40) seconds.
  - Storage Space is (50) KB.

Each small text of association rules of each site is sent to controller site

- Implantation of (PPAR-CC) Algorithm

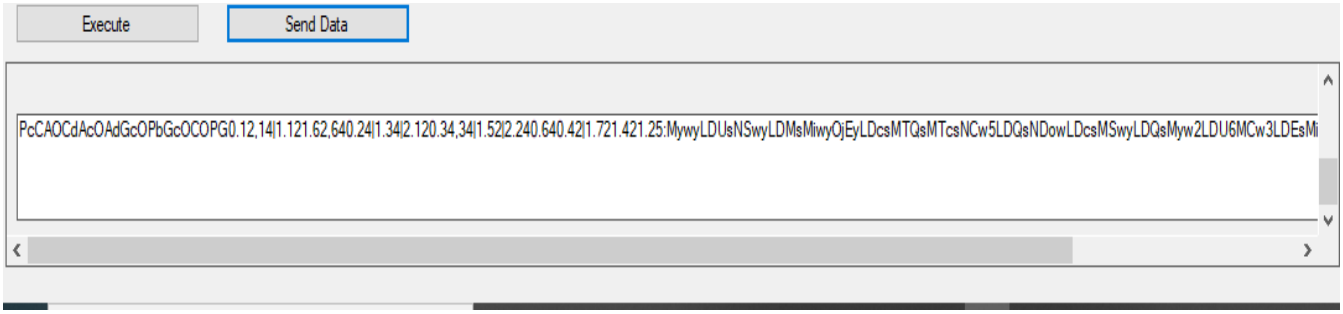


Figure (4-7) Change locations randomly then convert one record into incomprehensible small text of site 2

- ***Site2 (PPAR-CC) Algorithm implementation***
  - The output is very small and unintelligible text (only one line).
  - Execution time is less than (1) second.
  - Storage Space is (11) KB.

- Implementation for all system in controller site

Association Rules for controller site are explained as follows:

Table (4-17) represents the implementation of proposed system in controller site. As explained in proposed system steps (phase two) in chapter three, as a result of the implementation in controller site, Association Rules are extracted for all sites.

- **For all Datasets (1&2) of sites:**
  - Number of association rules is (23).
  - Execution time is less than (1) second.
  - Storage Space is (10) KB.

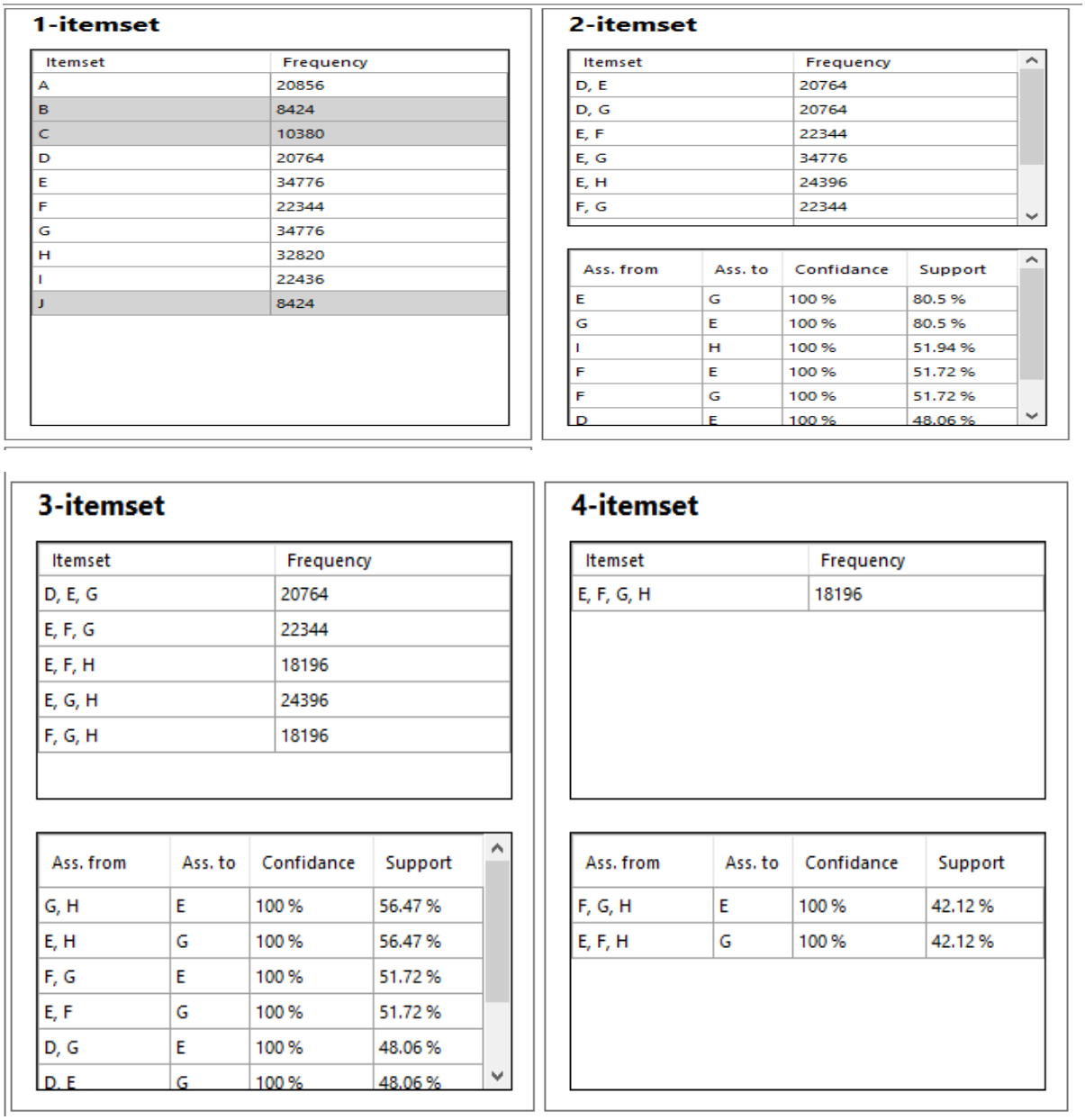
Implementation for all system in controller site

Table (4-17) Association rules of all system

No	K-itemset	K+1itemset	Ass. from	Ass. to
1	D	DE	D	E
2	D	DG	D	G
3	E	EF	E	F
4	E	EG	E	G
5	F	FG	F	G
6	F	FH	F	H
7	G	EG	G	E
8	I	HI	I	H
9	DE	DEG	DE	G
10	DG	DEG	DG	E
11	EF	EFG	EF	G
11	EF	EFG	EF	G
12	EF	EFH	EF	H
13	EH	EGH	EH	G
14	EH	EHI	EH	I
15	FG	EFG	FG	E
16	FG	FGH	FG	H
17	FH	FHI	FH	I
18	GH	EGH	GH	E
19	IJ	HIJ	IJ	H
20	AIJ	AHIJ	AIJ	H
21	EFH	EFGH	EFH	G
22	FGH	EFGH	FGH	E
23	GIJ	GHIJ	GIJ	H

4.5.2 Implementation standard Algorithms

1- A-priori Algorithm



1-itemset

Itemset	Frequency
A	20856
B	8424
C	10380
D	20764
E	34776
F	22344
G	34776
H	32820
I	22436
J	8424

2-itemset

Itemset	Frequency
D, E	20764
D, G	20764
E, F	22344
E, G	34776
E, H	24396
F, G	22344

Ass. from	Ass. to	Confidance	Support
E	G	100 %	80.5 %
G	E	100 %	80.5 %
I	H	100 %	51.94 %
F	E	100 %	51.72 %
F	G	100 %	51.72 %
D	E	100 %	48.06 %

3-itemset

Itemset	Frequency
D, E, G	20764
E, F, G	22344
E, F, H	18196
E, G, H	24396
F, G, H	18196

Ass. from	Ass. to	Confidance	Support
G, H	E	100 %	56.47 %
E, H	G	100 %	56.47 %
F, G	E	100 %	51.72 %
E, F	G	100 %	51.72 %
D, G	E	100 %	48.06 %
D, E	G	100 %	48.06 %

4-itemset

Itemset	Frequency
E, F, G, H	18196

Ass. from	Ass. to	Confidance	Support
F, G, H	E	100 %	42.12 %
E, F, H	G	100 %	42.12 %

Figure (4-8) A-priori Algorithm implementation of site 2

- ***Site2 A-priori Algorithm implementation***
  - Number of association rules only is (17).
  - Execution time is (8) minute.

- Storage Space is (600) KB.

2- A novel technique of Privacy Preserving Association Rule Mining

Suppose there is an example of 20 records of association rules as in Table (4-18) that will be passed in stochastic standard map, and the result of stochastic standard map will be explaind in Table (4-19).

Table (4-18) example of 20 recored of association rules

No.	Association from	Association to
R1	A, B, C, H	L, M, N, O, Q
R2	B, C, D, H	L, M, N, O, R
R3	A, B, C, H	N, O, P, R
R4	B, C, H	M, N, R
R5	A, B	D
R6	A, B	C, D
R7	B, M	N, O
R8	B, N, O	P, R,S
R9	B	H
R10	B	J
R11	A, B, C	H, J
R12	A, B	D, J
R13	B, H	L, M, N
R14	B, J	M, N, O, P
R15	A, B, C	H, J,M, N, O, P
R16	A, B	D,J ,L, M, N
R17	M, N, O, P	H, J
R18	A, B, C	M, N, O, P
R19	L, M, N	D, J
R20	A, C	D, G

Table (4-19) output of stochastic standard map on 20 rules

RULES	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
R1	A	B	C					H				L	M	N	O		Q									
R2		B	C	D				H				L	M	N	O			R								
R3	A	B	C					H						N	O	P		R								
R4		B	C					H					M	N				R								
R5	A	B		D																						
R6	A	B	C	D																						
R7		B											M	N	O											
R8		B												N	O	P		R	S							
R9		B						H																		
R10		B								J																
R11	A	B	C					H		J																
R12	A	B		D						J																
R13		B						H				L	M	N												
R14		B								J			M	N	O	P										
R15	A	B	C					H		J			M	N	O	P										
R16	A	B		D						J		L	M	N												
R17								H		J			M	N	O	P										
R18	A	B	C										M	N		P	Q	R								
R19				D						J		L	M	N												
R20	A		C	D			G																			

- ***implementation of stochastic standard map on 20 association rules***
  - The output is large Table that is larger than the size of original Table of association rules (It is four times the size of the original Table of association rules).
  - Execution time is (1) minute.
  - Storage Space is (100) KB.

4.6 Discussion

1- The proposed system (EPPAR-CC) is like any distributed system, as it contains a number of separate sites (site 1, site 2, …. site n), each of which works with sensor data streams. Different sites mine the data at the same time. This gives high performance in terms of mining time and storage space.

2- The proposed system (EPPAR-CC) consists of several Algorithms, including a proposed Algorithm for data mining ((DST-B-ARM) Algorithm for

association rules mining) for each site from the distributed system, as well as Algorithms to protect and hide the extracted knowledge((SARP-CC) & (PPAR-CC) Algorithms) and finally Algorithms to extract global knowledge (global association rules) for all sites.

3- In the proposed system, the proposed association rules mining (DST-B-ARM) Algorithm deals with sensors data streams, not static data, to reduce the storage space on the one hand, but dealing with data with dynamic properties also adds some complexity, as the data can only be scanned once.

4- In the proposed system, two different Algorithms ((SARP-CC) & (PPAR-CC) Algorithms) were used to protect the extracted knowledge, but they have the same basic structure. Here, the compression is not in the sense of compression to reduce the volume of data significantly, but it is considered the first step to change the shape of knowledge to an unknown form, so it is considered the first step of hiding.

5- The results of classical data streams mining Algorithms are considered speculative or discretionary and therefore because the data streams depending on the time factor that is not taken into consideration in the data mining process. In the proposed Algorithm (DST-B-ARM), the time factor is taken into consideration and more than that it is included in the basic steps in mining where we finally get strong association rules in addition to association time, so we transformed the extracted knowledge for data streams from estimated into exact through certain times.

6- The local mining of data at each site gives high performance better than sending huge amounts of data from each site to the controller site for once mining because in the second case it needs a very large storage space and also needs a large time to send and implement it. In addition to that, there are some other problems such as Sending large data via any network may cause some

data to be lost during transmission, and maintaining the security of large data is also more expensive computationally and even in terms of transmission time. In addition, we may lose some of the association rules that are important for certain sites.

7- Local data mining at each site and sending the resulted knowledge of each site to the controller site to be after that extracting the global knowledge removes the problem of false negative because any association rule within the global association rules is originally found within the association rules for one or more of distributed sites.

8- In the proposed system, the use of more than one Algorithm ((SARP-CC) & (PPAR-CC) Algorithms) to protect and conceal the extracted knowledge in different sites gives better performance in terms of security and privacy, because in order to uncover global knowledge (global association rules) of all system, it requires revealing the knowledge in all sites.

9- In the privacy preserving data mining Algorithms based on cryptography, we do not go into the details of the data mining process, but only the results from mining, while in the proposed system we take the data from the source (sensors) and perform the mining process using a proposed Algorithm and after obtaining the mining results we perform conservation operations Privacy of the results so that the knowledge extracted can only be disclosed to authorized persons.

4.7 Discussion of Comparisons and Accuracy of Results:

The proposed system was implemented to find the association rules for all sites in safe way which are more accurate than the global association rules which were found from all of the raw data by using traditional techniques, since proposed system guarantees correct and independent analysis for each site. (because it’s keeping the

private data at each site and mines its association rules which are computed at its own site) and then the association rules for all sides are mined from it.

1- Apriori Algorithm versus DST-B-ARM Algorithm

- Apriori Algorithm suffers from several challenges:-

1- Multiple scans are used on the original datasets to obtain the results.

2- Do not deal with the time factor and neglect it, if any.

3- Do not deal with dynamic data (data streams).

4- It takes a lot of time to implementation.

5- It is required that the data be stored before processing.

- These challenges were solved with the proposed Algorithm (DST-B-ARM) through:-

1- Use only one scan on the original datasets.

2- It deals with the time factor and includes it in the main steps of the Algorithm.

3- The time to implementation is relatively short.

4- It is not necessary to store data before processing.

5- Dealing with static & dynamic datasets.

2- The stochastic standard map technique versus the proposed techniques (SARP-CC & PPAR-CC)

- The stochastic standard map technique suffers from several challenges.

1- It needs a large storage space.

2- The execution time is relatively large compared to the amount of original knowledge.

3- It suffers from complexity because it deals with large Tables.

4- The resulting Table is twice the area of the original knowledge.

5- the knowledge pass through Large Tables, each Table represents a stage.

- These challenges are solved in the two proposed techniques (SARP-CC & PPAR-CC).

1- It needs a very small storage space.

2- It takes very little execution time.

3- Less complex and difficult to reveal original knowledge.

4- The final output is less space than the original knowledge.

5- The knowledge passes through four stages to reach the final output, and each stage has several steps.

3- Privacy

1- In order to preserve privacy, the proposed system contains two Algorithms to hide the extracted association rules (SARP-CC Algorithm & PPAR-CC Algorithm).

2- Each Algorithm hides the extracted association rules during four stages, this gives greater privacy and increases the difficulty in revealing the extracted association rules to unauthorized persons.

3- Both proposed Algorithms to preserve privacy produce a small text that is incomprehensible to unauthorized persons. This also increases the privacy of the mining output.

4- Some security standards were used to measure the strength of the proposed Algorithms as in Table (4-20), and the results were as follows

Table (4-20) NIST results of (SARP-CC Algorithm & PPAR-CC Algorithm)

Frequency analysis	0.0684
Block Frequency analysis	0.999
Cumulative Sums analysis	0.1306
Runs analysis	0.077
Longest Run analysis	0.0102

4- Storage Cost:

1- In each site, each record in sensor data streams of proposed system is scanned only once without need to store any record. This greatly reduces the storage space required.

2- The proposed system works with small text only (which is basically association rules record for each site) instead of huge quantity of records. Therefore, the proposed system will reduce required storage sized.

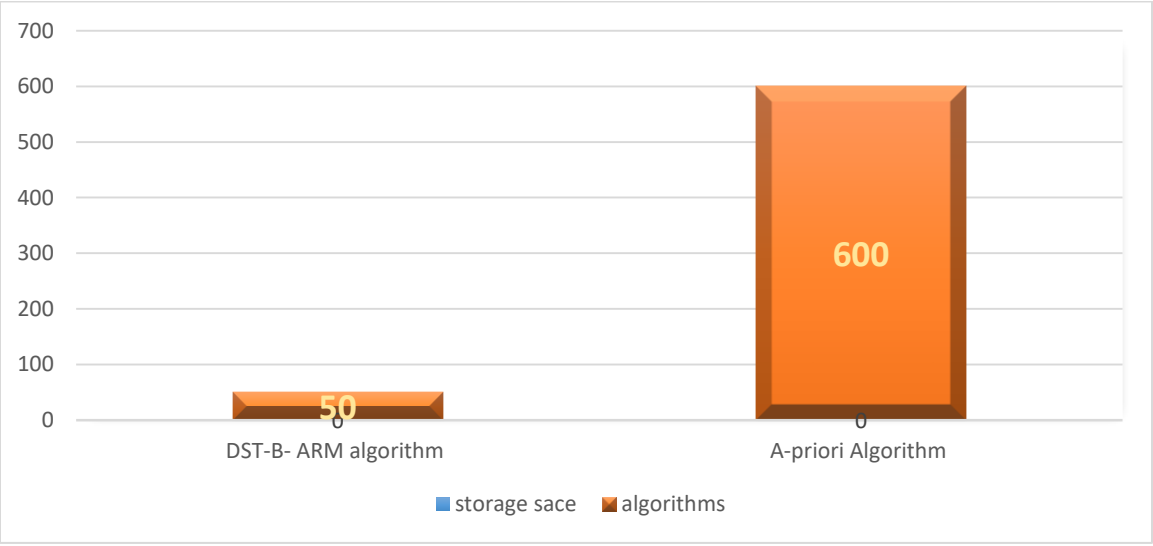


Figure (4-9) Comparison of storage space for DST-B-ARM Algorithm & A-priori Algorithm

5- Communication Cost:

Transferring a huge volume of data over network might take much time and also requires cost. The proposed system saves time and cost needed because it works on the distributed association rules (small text only) of each site instead of using the raw data of all sites.

6- Execution Time:

The proposed system needs less execution time because it used one scan only on sensor readings as well as, it works with association rules instead of all raw data. In other words, the proposed system works with small text only (which is basically association rules records for each site) instead of huge quantity of records.

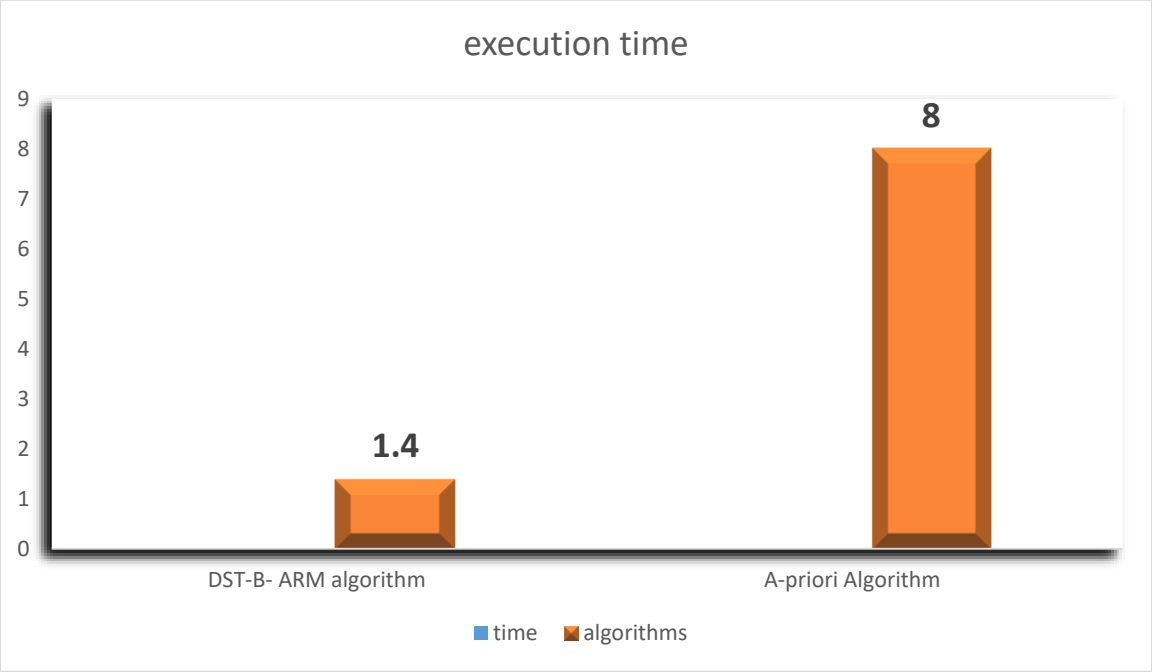


Figure (4-10) Comparison of execution time for DST-B-ARM Algorithm & A-priori Algorithm

7- Accuracy of results

1- With regard to the proposed association rules mining Algorithm (DST-B-ARM), the accuracy of its results was measured based on the results of the standard Algorithm (A-priori Algorithm) when applied to the same datasets, and it proved that the results are identical 100%.

2- With regard to the two methods of compression and coding, they also proved the accuracy of the results after decoding and obtaining the final results of the system (global association rules) and it proved that the results are identical 100% and there is no loss of association rules or the presence of fake association rules.

Chapter five

Conclusion and suggestions for future work

5.1 Conclusion

The most important conclusions that we will include and discuss in this section

1- In the proposed system, the size of the required storage space is reduced to the maximum extent in all the proposed Algorithms presented and this is considered one of the most important features mentioned to improve performance in modern research.

   A- Where we start first with the association rule mining Algorithm ((DST-B-ARM) Algorithm) that reduces the required storage space by dealing with data streams without need to store any raw data streams and extracting the required knowledge with the least storage space.

   B- In the Algorithms ((SARP-CC) & (PPAR-CC) Algorithms) to protect the association rules resulting from the mining Algorithm, where the size of information reduces from several records into only one record then Incomprehensibly small text that contains all the association rules for a specific site within the distributed environment.

   C- Finally, in the stage of extracting global knowledge at the controller site, where the knowledge (association rules) resulting from each site in the Incomprehensibly small text from several sites is unified in one record only that contains the knowledge (global association rules) extracted from all distributed sites.

2- The proposed system reduces time as much as possible because the time factor is one of the most important features to improve performance in modern research.

A- Time reduction begins with the data mining Algorithm, where the Algorithm works in all sites in the same time, i.e. at the same time, where huge amounts of data are mined in a relatively short time, and this gives improved performance.

B- There is also a time reduction in the (SARP-CC) & (PPAR-CC) Algorithms, where the amount of extracted knowledge is reduced to only one record, and then that record is easily converted into Incomprehensibly small text in a very short time, due to its small size.

C- In the process of sending an Incomprehensibly small text of knowledge from several sites distributed into the controller site, there is also a reduction in transmission times, because the smaller the volume of data, the less and easier the transmission time.

3- The proposed system improves the security and privacy of the knowledge resulting (association rules) from the mining process by: -

A- Using more than one Algorithm ((SARP-CC) & (PPAR-CC) Algorithms) to protect the knowledge in the different distributed sites to protect the knowledge generated at each site, because the detection of global knowledge by unauthorized persons and rival opponents requires disclosure of the knowledge resulting from each site in the distributed system.

B- The SARP-CC & PPAR-CC Algorithms pass the knowledge extracted at each site through four stages through which the form

of knowledge is changed from several records to a very small text that is not comprehensible, so unauthorized persons must expose all four stages to reach the knowledge extracted from one site, and all four stages of the two Algorithms must be exposed in order to gain access to the knowledge of the distributed system as a whole, this is almost impossible.

4- The data streams mining Algorithm ((DST-B-ARM) Algorithm) in the proposed system works incrementally depending on time, as well as we can control its implementation, i.e. it starts working from the specified time and ends or stops at the specified time as needed.

5- In the proposed system, a streaming data mining Algorithm ((DST-B-ARM) Algorithm) was used that depends on the time factor within the basic steps in it and thus it produces strong association rules with association time. In this case, the results of data streams mining process converted from an estimation that neglects the time factor into exact results specified by times representing the association time.

6- DST-B-ARM Algorithm of association rules mining finds the association rules in addition to the times of association and this is very important for many applications, including smart environments that depend on sensors, where finding the associative relations among device and other devices within specific periods of time provides more accurate information in order to make better decisions, as knowing the Correlational relationship without specifying the period of association does not provide accurate knowledge in order to make useful decisions.

5.2 Suggestions for future work

There are some suggestions for new researches which improve or develop this system:

1. Apply proposed system to other fields like web services data mining, medical diagnosis information system, mobile networks, chemical reactions, financial and economic data mining.
2. It would be interesting for future work to deal with new types of association rules such as weighted, fuzzy, rare, negative & multi-level association rules which are useful in some applications.
3. We can develop encryption methods and make them more powerful to provide more protection for the extracted knowledge.

References

[1] Meenakshi Bansa, Dinesh Grover and Dhiraj Sharma," Sensitivity Association Rule Mining using Weight based Fuzzy Logic",Global Journal of Enterprise Information System, 2017.

[2] M. Supriyamenon and Dr. P.Rajarajeswari Research Scholar," A Review on Association Rule Mining Techniques with Respect to their Privacy Preserving Capabilities", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 24 (2017).

[3] Madhuree Thakar, Shreya Dholaria and Ashutosh Abhangi, "Association Rule Hiding Technique - A Review", International Journal of Innovative Research in Computer and Communication Engineering, 2017.

[4] Jure Leskovec, Stanford Univ., Anand Rajaraman ,Milliway Labs, Jeffrey D. Ullman, Stanford Univ., “Mining of Massive Datasets” Copyright 2010, 2011, 2012, 2013, 2014 Anand Rajaraman, Jure Leskovec, and Jeffrey D. Ullman.

[5] Yousra Abdul Alsahib S. Aldeen, Mazleena Salleh and Mohammad Abdur Razzaque," A comprehensive review on privacy preserving data mining" Aldeen et al. SpringerPlus, 2015.

[6] Rana Saad Mohammed, Enas Mohammed Hussien & Jinan Redha Mutter “A novel technique of Privacy Preserving Association Rule Mining”, Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA) – IRAQ (9-10) May, 2016.

[7] Masooda Modaka and Rizwana Shaikhb, "Privacy Preserving Distributed Association Rule Hiding Using Concept Hierarchy" , 7th International Conference on Communication, Computing and Virtualization 2016.

[8] Naadiya Khuda Bux, Mingming Lu, Jianxin Wang , Saajid Hussain and Yazan Aljeroudi," Efficient Association Rules Hiding Using Genetic Algorithms",

Symmetry 2018, 10, 576; doi:10.3390/sym10110576 www.mdpi.com/journal/symmetry.

[9] RICARDO MENDES AND JOAO P. VILELA," Privacy-Preserving Data Mining: Methods, Metrics, and Applications", IEEE. Translations and content mining are permitted for academic research only, 2018.

[10] Surendra H and Dr. Mohan H S "Preserving Privacy of Sensitive Itemsets using Controlled Perturbation of Closed Itemsets", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 5 (2019) pp. 1177-1185.

[11] Salha Albehairi," A Privacy-Preserving Framework for Collaborative Association Rule Mining in Cloud", Master thesis, Montclair State University, May 2019.

[12] Ma´rcio Alencar, Raimundo Barreto, Hora´cio Fernandes, Eduardo Souto and Richard Pazzi," DARE: A decentralized association rules extraction scheme for embedded data sets in distributed IoT devices", International Journal of Distributed Sensor Networks, 2020, Vol. 16(10).

[13] Venkatesh Kumar. M & Dr. C. Lakshmi, "AN EFFICIENT SECURE COMPUTATION FOR PRIVACY PRESERVING DATA MINING IN MULTI PARTY COMPUTATION (MPC) – A REVIEW", International Journal of Advanced Research in Engineering and Technology, 11(10), 2020.

[14] Nikunj Domadiya and Udai Pratap Rao, "Privacy Preserving Association Rule Mining on Distributed Healthcare Data: COVID-19 and Breast Cancer Case Study", Department of Computer Engineering, Sardar Vallabhbhai National Institute of Technology, Surat 395007, India, SN Computer Science (2021) 2:418.

[15] Anvita Srivastava, "Privacy Preserving Data Mining in Electronic Health Record using K-anonymity and Decision Tree", Anvita Srivastava et al. /

International Journal of Computer Science & Engineering Technology (IJCSET), ISSN : 2229-3345 Vol. 6 No. 07 Jul 2015.

[16] Sam Fletcher," Data Mining and Privacy: Modeling Sensitive Data with Differential Privacy", A thesis of Doctorate of Philosophy, School of Computing and Mathematics June 2017.

[17] M.Chalapathi Rao1, A.Kiran Kumar2," Challenges arise of Privacy Preserving Big Data Mining Techniques", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 04 Issue: 05 | May -2017.

[18] Gayathiri. P, Dr. B Poorna," Association Rule Hiding for Privacy Preserving Data Mining: A Survey on Algorithmic Classifications", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 23 (2017) pp. 13917-13926.

[19] Karishma Chopda1 , Apurva Rote2 , Komal Gaikwad3 , Priyanka Gachale," Association Rule Mining Method for Applying Encryption Techniques in Transaction Data", International Journal of Engineering Science and Computing, May 2017.

[20] M. Supriyamenon and Dr. P.Rajarajeswari, "A Review on Association Rule Mining Techniques with Respect to their Privacy Preserving Capabilities", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 24 (2017) pp.

[21] Omoyele, Tobi Deborah and Akinola, Solomon Olalekan," AN IMPROVED ALGORITHM FOR PRIVACY PRESERVING QUANTITATIVE DATA USING ASSOCIATION RULE MINING AND PERTURBATION TECHNIQUE", GESJ: Computer Science and Telecommunications 2017|No.1(51).

[22] Madhuree Thakar1 , Shreya Dholaria2 , Ashutosh Abhangi3," Association Rule Hiding Technique - A Review", International Journal of Innovative Research in

Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Website: www.ijircce.com Vol. 5, Issue 2, February 2017.

[23] A. Damodar, C.Rajeev, M.Srinivas Reddy," Privacy Preserving in Data Mining with No Data Loss with a Combinational Scheme", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4S3, December 2019.

[24] Farsad Zamani Boroujeni and Doryaneh Hossien Afshari," An Efficient Rule-Hiding Method For Privacy Preserving in Transactional Databases", CIT. Journal of Computing and Information Technology, Vol. 25, No. 4, December 2017, 279–290 279 doi: 10.20532/cit.2017.1003680.

[25] Shilpa Rathod, Dr. Darshana Patel, " Survey on Privacy Preserving Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV9IS060568 (This work is licensed under a Creative Commons Attribution 4.0 International License.) Published by : www.ijert.org Vol. 9 Issue 06, June-2020.

[26] Elisa Bertino, Igor Nai Fovino & Loredana Parasiliti Provenza,” A Framework for Evaluating Privacy Preserving Data Mining Algorithms*”, Data Mining and Knowledge Discovery, 11, 121–154, Springer Science+Business Media, Inc. DOI: 10.1007/s10618-005-0006-6, 2005.

[27] " Streaming Data vs. Static Data", Copyright © SAS Institute Inc. All Rights Reserved. Last updated: December 12, 2018.

[28] Gayathri Devi N Manikandan K," Improved perturbation technique privacy-preserving rotation-based condensation Algorithm for privacy preserving in big data stream using Internet of Things", © 2020 John Wiley & Sons, Ltd.

[29] Sanket P. Modi1 Ashil R. Patel2," Privacy Preserving Data Stream Mining using Two Phase Geometric Data Perturbation", IJSRD - International Journal for

Scientific Research & Development| Vol. 3, Issue 01, 2015 | ISSN (online): 2321-0613.

[30] Ankit Jasoliya1 , Tejal Patel2, " A Survey: Privacy Preserving Techniques in Data Stream Mining", INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD ISSN – 2455-0620 Volume - 3, Issue - 5, May – 2017.

[31] Ankit Jasoliya1 , Tejal Patel," A Survey: Privacy Preserving Techniques in Data Stream Mining", INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD ISSN – 2455-0620 Volume - 3, Issue - 5, May – 2017.

[32] Paresh Solanki1 , Sanjay Garg2 , Hitesh Chhinkaniwala3," Privacy Preserving Data Stream Mining Using Hybrid Geometric Data Perturbation", IJRECE VOL. 5 ISSUE 3 JULY.-SEPT. 2017.

[33] Paresh Solanki1 , Sanjay Garg2 , Hitesh Chhinkaniwala3, "Privacy Preserving Data Stream Mining Using Hybrid Geometric Data Perturbation" , IJRECE VOL. 5 ISSUE 3 JULY.-SEPT. 2017 ISSN: 2393-9028 (PRINT) |ISSN: 2348-2281 (ONLINE).

[34] Richa Purohit & Deepshikha Bhargava, " An illustration to secured way of data mining using privacy preserving data mining", Journal of Discrete Mathematical Sciences and Cryptography, 20 Dec 2017.

[35] Wensheng Gan, Chun-Wei Jerry Lin, Han-Chieh Chao, Justin Zhan, “Data mining in distributed environment: a survey”, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 7(6):e1216, DOI: 10.1002/widm.1216, July 2017.

[36] V. Baby, N. Subhash Chandra, PhD, " Privacy-Preserving Distributed Data Mining Techniques: A Survey", International Journal of Computer Applications (0975 – 8887) Volume 143 – No.10, June 2016.

[37] Yousra Abdul Alsahib S. Aldeen1,2*, Mazleena Salleh1 and Mohammad Abdur Razzaque1, " A comprehensive review on privacy preserving data mining", Aldeen et al. Springer-Plus (2015) 4:694.

[38] Jun Liu, Yuan Tian, Yu Zhou, Yang Xiao, Nirwan Ansari, " Privacy preserving distributed data mining based on secure multi-party computation", Computer Communications, Volume 153, 1 March 2020, Pages 208-216.

[39] S. Hariraman1, Dr. S. Velmurugan2, "An Enhanced Privacy Preserving Techniques for Asynchronous Streaming Data Mining in Distributed Environment", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, www.ijert.org ECLECTIC – 2020.

[40] Kurapati Praveena1*, Gudla Sirisha2 , Satukumati Suresh Babu3 , Panchala Sambasiva Rao, " Efficient Method in Association Rule Hiding for Privacy Preserving with Data Mining Approach", Ingenierie des Systemes d'Information Vol. 24, No. 1, February, 2019, pp. 47-50.

[41] N. Subhash Chandra, " Privacy preserving association rule mining based on homomorphic computations", International Journal of Information Privacy, Security and Integrity (IJIPSI), Vol. 3, No. 4, 2018.

[42] A. Anitha1, Dr.K. Selvam2, " PROTECTION PRESERVING DATA MINING (PPDM) METHOD FOR CROSS DIVIDED DATA", Journal of Critical Reviews ISSN- 2394-5125 Vol 7, Issue 11, 2020.

[43] Praveen Kumar Gopagoni, Mohan Rao S K," Distributed elephant herding optimization for grid-based privacy association rule mining", Data Technologies and Applications, ISSN: 2514-9288, 15 May 2020.

[44] Wensheng Gan, Jerry Chun, Wei Lin, Han-Chieh Chao, Justin Zhan, “Data mining in distributed environment: a survey”, WIREs Data Mining and Knowledge Discovery, 18 July 2017.

[45] Gayathiri P, Dr. B Poorna, " Association Rule Hiding Techniques for Privacy Preserving Data Mining: A Study", International Journal of Advanced Computer Science and Applications(IJACSA), Volume 6 Issue 12, 2015.

[46] Akbar Telikani, Amir H.Gandomi, Asadollah Shahbahrami and Mohammad Naderi Dehkordi, " Privacy-preserving in association rule mining using an improved discrete binary artificial bee colony", Expert Systems with Applications, Volume 144, 15 April 2020.

[47] Harendra Chahar, B N Keshavamurthy & Chirag Modi, “Privacy-preserving distributed mining of association rules using Elliptic-curve cryptosystem and Shamir’s secret sharing scheme”, Springer, DOI:10.1007/S12046-017-0743-4Corpus ID: 126194890, 17 November 2017.

[48] Hyeong-Jin Kim, Jae-Hwan Shin, Young-ho Song, Jae-Woo Chang, " Privacy-Preserving Association Rule Mining Algorithm for Encrypted Data in Cloud Computing", IEEE 29 August 2019, Conference Location: Milan, Italy.

[49] Radhika Garg et. Al, " Association Rule Mining Algorithms through Vertical and Horizontal Data Layouts: Implementation and Performance Comparison", International Journal of Advanced Science and Technology, Vol. 29 No. 2 (2020): Vol 29 No 2 (2020).

[50] LI Chengyan, Shixiang FENG & Guanglu SUN, "DCE -miner: an association rule mining Algorithm for multimedia based on the MapReduce framework", Springer Link, 07 June 2020.

[51] Hsiao-Kang Lin, Cheng-Huan Hsieh, Nai-Chieh Wei and Yi-Chun Peng, "Association rules mining in R for product performance management in industry 4.0", Procedia CIRP Volume 83, 2019, Pages 699-704.

[52] Elif Varol Altay and Bilal Alatas, " Intelligent optimization Algorithms for the problem of mining numerical association rules", Physica A: Statistical Mechanics and its Applications Volume 540, 15 February 2020, 123142.

[53] Diana-Lucia Miholca, Gabriela Czibula, Liana Maria Crivei, "A new incremental relational association rules mining approach", Procedia Computer Science, Volume 126, 2018, Pages 126-135.

[54] Iyad Aqra 1,* ,† , Norjihan Abdul Ghani 1,* ,† , Carsten Maple 2,3,‡, José Machado 4,‡ and Nader Sohrabi Safa, " Incremental Algorithm for Association Rule Mining under Dynamic Threshold", Appl. Sci. 2019, 9, 5398; doi:10.3390/app9245398.

[55] P. Naresh ; R. Suguna, " Association Rule Mining Algorithms on Large and Small Datasets: A Comparative Study", IEEE, 16 April 2020.

[56] Chin-Hoong Chee, Jafreezal Jaafar, Izzatdin Abdul Aziz, Mohd Hilmi Hasan & William Yeoh," Algorithms for frequent itemset mining: a literature Review", Artif Intell Rev (2019) 52:2603–2621 https://doi.org/10.1007/s10462-018-9629-z, spring, 2019.

[57] Sudhir Tirumalasetty, Aruna Jadda and Sreenivasa Reddy Edara," An Enhanced Apriori Algorithm for Discovering Frequent Patterns with Optimal Number of Scans", India, 2015.

[58] Hong-Jun Jang, Yeongwook Yang, Ji Su Park and Byoungwook Kim," FP-Growth Algorithm for Discovering Region-Based Association Rule in the IoT Environment", Electronics 2021, 10, 3091.

[59] Manjit kaur and Urvashi Grag," ECLAT Algorithm for Frequent Itemsets Generation", International Journal of Computer Systems (ISSN: 2394-1065), Volume 01– Issue 03, December, 2014.

[60] Emad Kadum Jabbar, "New Algorithms for Discovering Association Rules", PHD, thesis, Department of Computer Science of University of Technology, 2005.

[61] Prakhar Shukla, Parnab Kumar Chanda, Ramnath Jayachandran, Ashok Subash, "A Framework for User Routine Discovery in Smart Homes", 2018 IEEE 6th International Conference on Future Internet of Things and Cloud.

[62] AMAL ALSAEH," PARALLEL ASSOCIATION RULE MINING ON SEMANTIC AND BIG IOT DATA", M.Sc., Department of Computer Engineering, July 2018.

[63] Ma´rcio Alencar, Raimundo Barreto , Hora´cio Fernandes, Eduardo Souto and Richard Pazzi, "DARE: A decentralized association rules extraction scheme for embedded data sets in distributed IoT devices", International Journal of Distributed Sensor Networks 2020, Vol. 16(10).

[64] Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang, "Data Mining for Internet of Things: A Survey", IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 16, NO. 1, FIRST QUARTER 2014.

[65] Ms. Vinaya Sawanta *, Dr. Ketan Shah, " Performance Evaluation of Distributed Association Rule Mining Algorithms", Vinaya Sawant and Ketan Shah / Procedia Computer Science 79 (2016) 127 – 134.

[66] R. Subha, "P-tree oriented association rule mining of multiple data sources", International Journal of Enterprise Network Management (IJENM), Vol. 10, No. 3/4, 2019.

[67] Khushbu Agrawal1 and Vandan Tewari2, "Privacy-Preservation in Collaborative Association Rule Mining for Outsourced Data", International Journal of Distributed and Cloud Computing Volume 5 Issue 2 December 2017.

[68] Pritam Chavan, Rahul Purushothaman, Satish, Aditya Anasane and Guide Gayatri Hegde, "Proposed Secure Mining of Association Rules in Horizontally Distributed Databases", DEPARTMENT OF INFORMATION TECHNOLOGY 2017.

[69] K. S. Ranjith, Yang Zhenning, Ronnie D. Caytiles* and N. Ch. S. N. Iyengar, "Comparative Analysis of Association Rule Mining Algorithms for the Distributed Data", International Journal of Advanced Science and Technology Vol.102 (2017), pp.49-60.

[70] Vadlana Baby, Dr. N. Subhash Chandra, "Privacy Preserving Distributed Association Rule Mining Algorithm for Vertically Partitioned Data", International Journal of Computer Science and Information Security (IJCSIS), Vol. 15, No. 8, Augus 2017.

[71] Rana, S., Santhi Thilagam, P., " Hierarchical homomorphic encryption based privacy preserving distributed association rule mining", 13th International Conference on Information Technology, ICIT 2014, 2014, Vol., pp.379-385.

[72] NIKUNJ DOMADIYA* and UDAI PRATAP RAO, " Privacy-preserving association rule mining for horizontally partitioned healthcare data: a case study on the heart diseases", Indian Academy of Sciences, Sådhanå (2018) 43:127.

[73] 1P. R. S. Naidu, 2B. Prasanth Kumar, 3Dr. P. Sateesh, 4Dr. B. Srinivas, 5Chandra Sekhar Darapaneni, "A Novel Approach of Association Rule Hiding Using DBCT (Distortion, Blocking and Cryptographic Technique)", Helix Vol. 9 (1): 4781- 4790, DOI 10.29042/2019-4781-4790, 2019.

[74] Heri Nurdiyanto1, Robbi Rahim2*, Ansari Saleh Ahmar3, Muhammad Syahril4, Muhammad Dahria5 and Herlina Ahmad6 “Secure a Transaction Activity with Base64 Algorithm and Word Auto Key Encryption Algorithm” 2nd International Conference on Statistics, Mathematics, Teaching, and Research, IOP Conf. Series: Journal of Physics: Conf. Series 1028 (2018) 012053 doi:10.1088/1742-6596/1028/1/012053.

[75] Săcăleanu, S.I., R. Stoian, D. M. Ofrim, “An adaptive Huffman Algorithm for data compression in wireless sensor networks” Proceedings of the International Symposium on Signals, Circuits and Systems, Jun. 30-Jul. 1, IEEE Xplore Press, Lasi, Romania, pp: 1-4. DOI: 10.1109/ISSCS.2011.5978764, 2011.

[76] Renugadevi, S. and P.S.N. Darisini,” Huffman and Lempel-Ziv based data compression Algorithms for wireless sensor networks”, Proceedings of the International Conference on Pattern Recognition, Informatics and Mobile

Engineering, Feb. 21-22, IEEE Xplore Press, Salem, India, pp: 461-463. DOI: 10.1109/ICPRIME.2013.6496521, 2013.

[77] Oswald, C. and B. Sivaselvan, “An optimal text compression Algorithm based on frequent pattern mining”, J. Ambient Intell. Human Comput., 9: 803- 822. DOI: 10.1007/s12652-017-0540-2, 2018.

[78] Oswald, C., A.I. Ghosh and B. Sivaselvan, “An efficient text compression Algorithm-data mining perspective”, Proceedings of the 3rd International Conference on Mining Intelligence and Knowledge Exploration, Dec. 09-11, Springer, Hyderabad, India, pp: 563-575. DOI: 10.1007/978-3-319-26832-3_53, 2015.

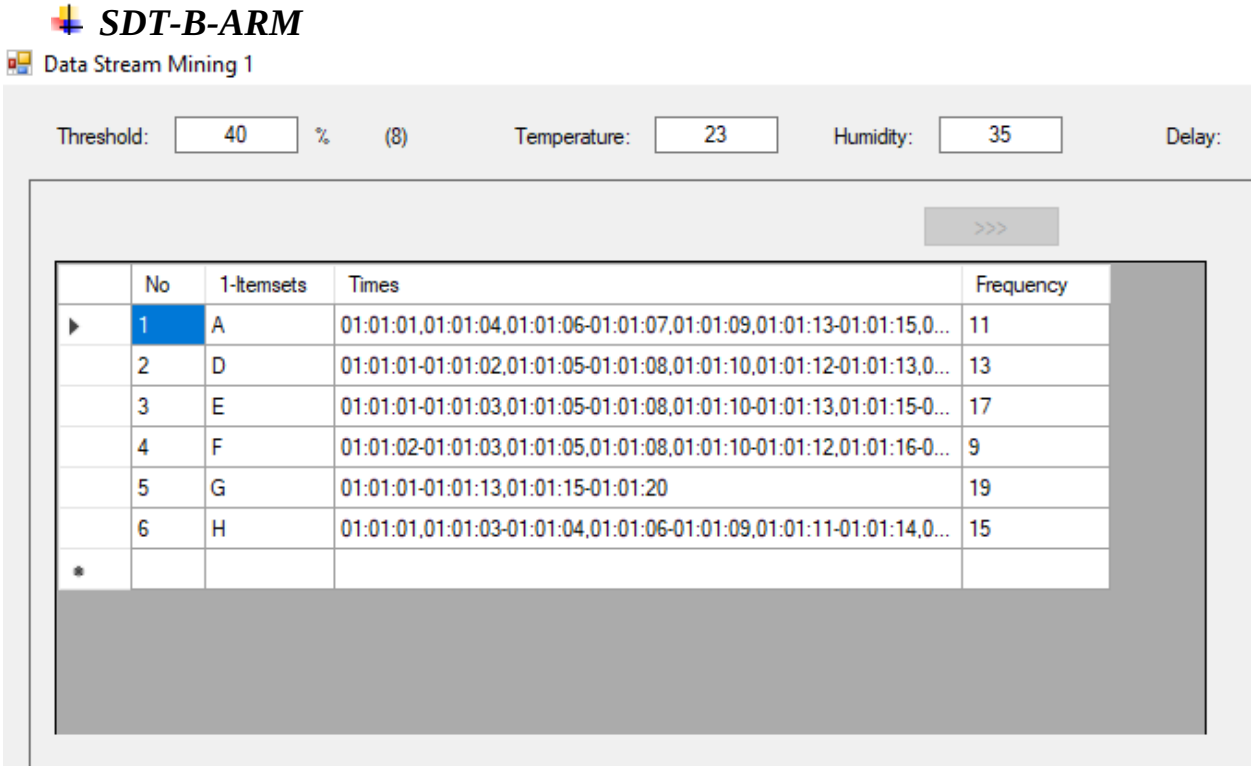
[79] Wang, W.J. and C.H. Lin,” Code compression for embedded systems using separated dictionaries”, IEEE Trans. Very Large Scale Integrat. Syst., 24: 266-275. DOI: 10.1109/TVLSI, 2015.

[80] Rajput, K.K., “Are Huffman trees balanced?”, https://www.quora.com/Are-Huffman-treesbalanced. 2018.

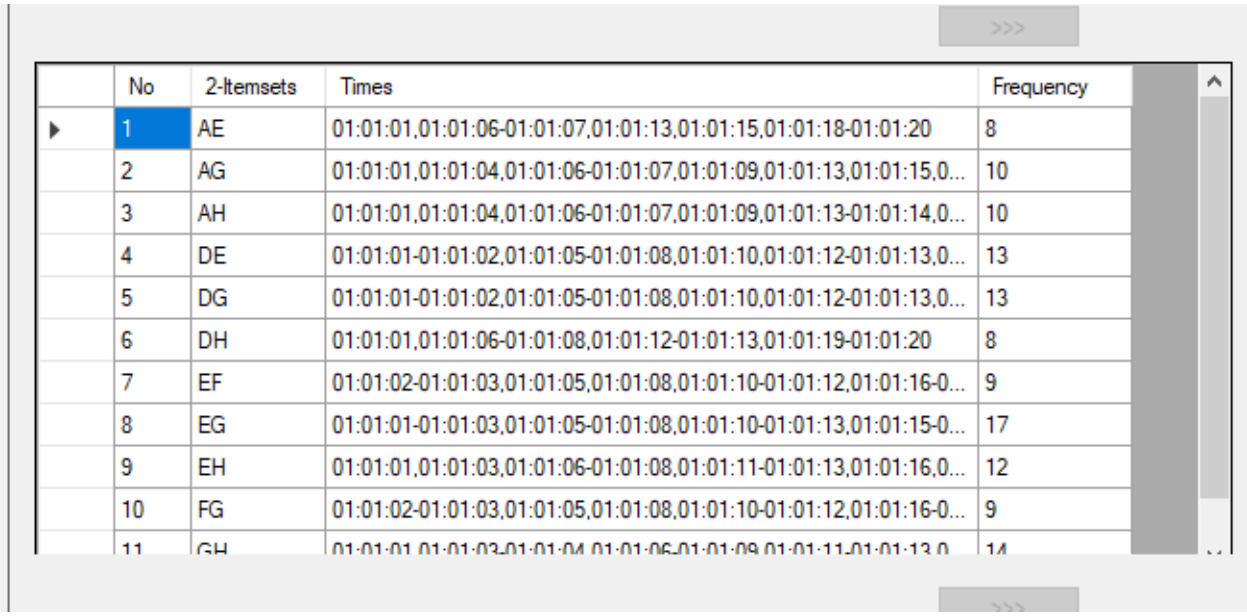
[81] Zaid Munthir Jawad Abdul Hadi, “Intelligent Internet of Everything Security Using Hyper-Chaotic System”, Doctor dissertation in Information Technology-Software, University of Babylon, 2021.

Appendix

tid	Sensors reading	time
1.	1,0,0,24,37	9:01
2.	0,0,1,24,37	9:02
3.	0,0,0,22,38	9:03
4.	1,0,0,21,33	9:04
5.	0,0,1,24,37	9:05
6.	1,0,0,24,37	9:06
7.	1,0,0,24,37	9:07
8.	0,0,0,23,38	9:08
9.	1,0,0,21,33	9:09
10.	0,0,1,24,37	9:10
11.	0,0,0,22,39	9:11
12.	0,0,0,24,40	9:12
13.	1,0,0,24,37	9:13
14.	1,1,0,21,33	9:14
15.	1,0,1,24,37	9:15
16.	0,0,0,22,36	9:16
17.	0,0,1,24,37	9:17
18.	1,0,0,22,39	9:18
19.	1,0,0,24,37	9:19
20.	1,0,0,25,36	9:20



	No	1-Itemsets	Times	Frequency
▶	1	A	01:01:01,01:01:04,01:01:06-01:01:07,01:01:09,01:01:13-01:01:15,0...	11
	2	D	01:01:01-01:01:02,01:01:05-01:01:08,01:01:10,01:01:12-01:01:13,0...	13
	3	E	01:01:01-01:01:03,01:01:05-01:01:08,01:01:10-01:01:13,01:01:15-0...	17
	4	F	01:01:02-01:01:03,01:01:05,01:01:08,01:01:10-01:01:12,01:01:16-0...	9
	5	G	01:01:01-01:01:13,01:01:15-01:01:20	19
	6	H	01:01:01,01:01:03-01:01:04,01:01:06-01:01:09,01:01:11-01:01:14,0...	15
*				



No	2-Itemsets	Times	Frequency
1	AE	01:01:01,01:01:06-01:01:07,01:01:13,01:01:15,01:01:18-01:01:20	8
2	AG	01:01:01,01:01:04,01:01:06-01:01:07,01:01:09,01:01:13,01:01:15,0...	10
3	AH	01:01:01,01:01:04,01:01:06-01:01:07,01:01:09,01:01:13-01:01:14,0...	10
4	DE	01:01:01-01:01:02,01:01:05-01:01:08,01:01:10,01:01:12-01:01:13,0...	13
5	DG	01:01:01-01:01:02,01:01:05-01:01:08,01:01:10,01:01:12-01:01:13,0...	13
6	DH	01:01:01,01:01:06-01:01:08,01:01:12-01:01:13,01:01:19-01:01:20	8
7	EF	01:01:02-01:01:03,01:01:05,01:01:08,01:01:10-01:01:12,01:01:16-0...	9
8	EG	01:01:01-01:01:03,01:01:05-01:01:08,01:01:10-01:01:13,01:01:15-0...	17
9	EH	01:01:01,01:01:03,01:01:06-01:01:08,01:01:11-01:01:13,01:01:16,0...	12
10	FG	01:01:02-01:01:03,01:01:05,01:01:08,01:01:10-01:01:12,01:01:16-0...	9
11	GH	01:01:01,01:01:03-01:01:04,01:01:06-01:01:09,01:01:11-01:01:13,0...	14

No	3-Itemsets	Times	Frequency
1	AEG	01:01:01,01:01:06-01:01:07,01:01:13,01:01:15,01:01:18-01:01:20	8
2	AGH	01:01:01,01:01:04,01:01:06-01:01:07,01:01:09,01:01:13,01:01:18-0...	9
3	DEG	01:01:01-01:01:02,01:01:05-01:01:08,01:01:10,01:01:12-01:01:13,0...	13
4	DEH	01:01:01,01:01:06-01:01:08,01:01:12-01:01:13,01:01:19-01:01:20	8
5	DGH	01:01:01,01:01:06-01:01:08,01:01:12-01:01:13,01:01:19-01:01:20	8
6	EFG	01:01:02-01:01:03,01:01:05,01:01:08,01:01:10-01:01:12,01:01:16-0...	9
7	EGH	01:01:01,01:01:03,01:01:06-01:01:08,01:01:11-01:01:13,01:01:16,0...	12
*			

B

No	4-Itemsets	Times	Frequency
1	DEGH	01:01:01,01:01:06-01:01:08,01:01:12-01:01:13,01:01:19-01:01:20	8

No	K-itemset	k-1itemset	Association time	Time code	Ass. from	Ass. to
1	EF	F	01:01:02-01:01:17	1	F	E
2	DE	D	01:01:01-01:01:20	2	D	E
3	DG	D	01:01:01-01:01:20	2	D	G
4	EG	E	01:01:01-01:01:20	2	E	G
5	FG	F	01:01:02-01:01:17	1	F	G
6	AEG	AE	01:01:01-01:01:20	2	AE	G
7	DEH	DH	01:01:01-01:01:20	2	DH	E
8	DEG	DE	01:01:01-01:01:20	2	DE	G
9	DEG	DG	01:01:01-01:01:20	2	DG	E
10	DGH	DH	01:01:01-01:01:20	2	DH	G
11	EGH	EH	01:01:01-01:01:20	2	EH	G

Execute Send Data

No	K-itemset	k-1itemset	Association time	Time code	Ass. from	Ass. to
7	DEH	DH	01:01:01-01:01:20	2	DH	E
8	DEG	DE	01:01:01-01:01:20	2	DE	G
9	DEG	DG	01:01:01-01:01:20	2	DG	E
10	DGH	DH	01:01:01-01:01:20	2	DH	G
11	EGH	EH	01:01:01-01:01:20	2	EH	G
12	EFG	EF	01:01:02-01:01:17	1	EF	G
13	EFG	FG	01:01:02-01:01:17	1	FG	E
14	DEGH	DEH	01:01:01-01:01:20	2	DEH	G
15	DEGH	DGH	01:01:01-01:01:20	2	DGH	E

C

- *A-priori*

tid	A	B	C	D	E	F	G	H	I	J
1	1	0	0	1	1	0	1	1	0	0
2	0	0	1	1	1	1	1	0	0	0
3	0	0	0	0	1	1	1	1	1	0
4	1	0	0	0	0	0	1	1	1	1
5	0	0	1	1	1	1	1	0	0	0
6	1	0	0	1	1	0	1	1	0	0
7	1	0	0	1	1	0	1	1	0	0
8	0	0	0	1	1	1	1	1	0	0
9	1	0	0	0	0	0	1	1	1	1
10	0	0	1	1	1	1	1	0	0	0
11	0	0	0	0	1	1	1	1	1	0
12	0	0	0	1	1	1	1	1	0	0
13	1	0	0	1	1	0	1	1	0	0
14	1	1	0	0	0	0	0	1	1	1
15	1	0	1	1	1	0	1	0	0	0
16	0	0	0	0	1	1	1	1	1	0
17	0	0	1	1	1	1	1	0	0	0
18	1	0	0	0	1	0	1	1	1	0
19	1	0	0	1	1	0	1	1	0	0
20	1	0	0	1	1	0	1	1	0	0

1-Itemset	Frequency
A	11
B	1
C	5
D	13
E	17
F	9
G	19
H	15
I	7
J	3

2-itemset	*Frequency*
AD	***5***
AE	***8***
AF	***0***
AG	***10***
AH	***10***
AI	***5***
DE	***13***
DF	***5***
DG	***13***
DH	***8***
DI	***0***
EF	***9***
EG	***17***
EH	***12***
EI	***4***
FG	***9***
FH	***5***
FI	***3***
GH	***14***

3-itemset	*Frequency*
AEG	*8*
AGH	*9*
DEG	*13*
DEH	*8*
DGH	*8*
EFG	*9*
EGH	*12*

1-itemset

Itemset	Frequency
A	11
B	1
C	5
D	13
E	17
F	9
G	19
H	15
I	7
J	3

2-itemset

Itemset	Frequency
A, E	8
A, G	10
A, H	10
D, E	13
D, G	13
D, H	8

Ass. from	Ass. to	Confidance	Support
E	G	100 %	85 %
D	E	100 %	65 %
D	G	100 %	65 %
F	E	100 %	45 %
F	G	100 %	45 %

3-itemset

Itemset	Frequency
A, G, H	9
D, E, G	13
D, E, H	8
D, G, H	8
E, F, G	9
E, G, H	12

Ass. from	Ass. to	Confidance	Support
D, G	E	100 %	65 %
D, E	G	100 %	65 %
E, H	G	100 %	60 %
F, G	E	100 %	45 %
E, F	G	100 %	45 %
A, E	G	100 %	40 %

4-itemset

Itemset	Frequency
D, E, G, H	8

Ass. from	Ass. to	Confidance	Support
D, G, H	E	100 %	40 %
D, E, H	G	100 %	40 %

A-priori Algorithm of applicable example

- ***Site2 A-priori Algorithm implementation***
  - Number of association rules only is (17).
  - Execution time is (8) minute.
  - Storage Space is (600) KB.

1-itemset

Itemset	Frequency
A	20856
B	8424
C	10380
D	20764
E	34776
F	22344
G	34776
H	32820
I	22436
J	8424

2-itemset

Itemset	Frequency
D, E	20764
D, G	20764
E, F	22344
E, G	34776
E, H	24396
F, G	22344

Ass. from	Ass. to	Confidance	Support
E	G	100 %	80.5 %
G	E	100 %	80.5 %
I	H	100 %	51.94 %
F	E	100 %	51.72 %
F	G	100 %	51.72 %
D	E	100 %	48.06 %

3-itemset

Itemset	Frequency
D, E, G	20764
E, F, G	22344
E, F, H	18196
E, G, H	24396
F, G, H	18196

Ass. from	Ass. to	Confidance	Support
G, H	E	100 %	56.47 %
E, H	G	100 %	56.47 %
F, G	E	100 %	51.72 %
E, F	G	100 %	51.72 %
D, G	E	100 %	48.06 %
D, E	G	100 %	48.06 %

4-itemset

Itemset	Frequency
E, F, G, H	18196

Ass. from	Ass. to	Confidance	Support
F, G, H	E	100 %	42.12 %
E, F, H	G	100 %	42.12 %

A novel technique of Privacy Preserving Association Rule Mining 2016: By using stochastic standard map

No.	Association from	Association to
R1	A, B, C, H	L, M, N, O, Q
R2	B, C, D, H	L, M, N, O, R
R3	A, B, C, H	N, O, P, R
R4	B, C, H	M, N, R
R5	A, B	D
R6	A, B	C, D
R7	B, M	N, O
R8	B, N, O	P, R,S
R9	B	H
R10	B	J
R11	A, B, C	H, J
R12	A, B	D, J
R13	B, H	L, M, N
R14	B, J	M, N, O, P
R15	A, B, C	H, J,M, N, O, P
R16	A, B	D,J ,L, M, N
R17	M, N, O, P	H, J
R18	A, B, C	M, N, O, P
R19	L, M, N	D, J
R20	A, C	D, G

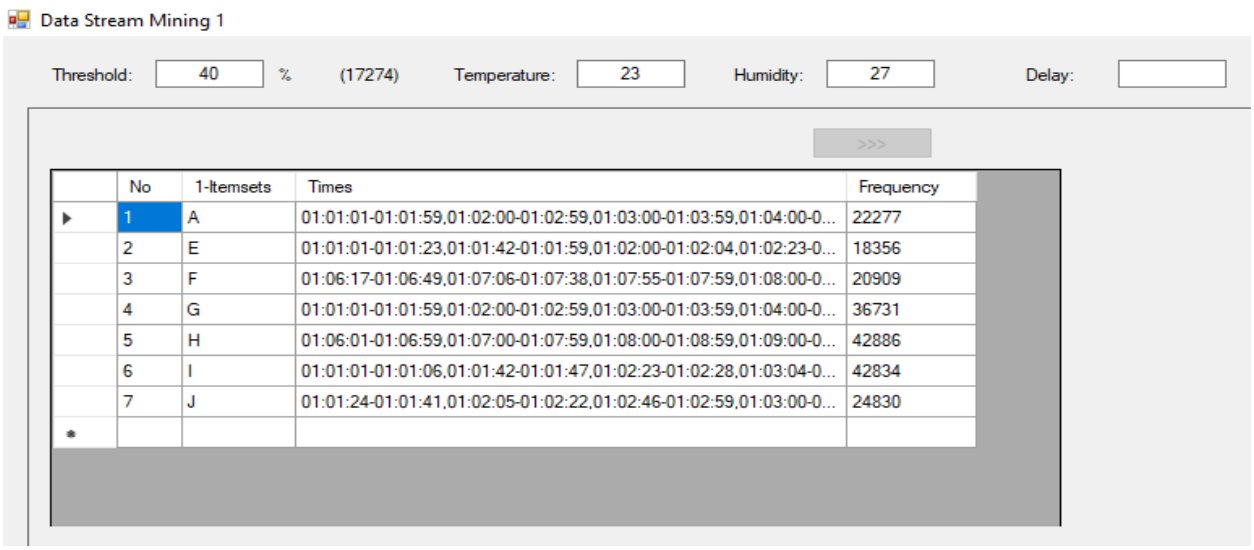
RULES	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
R1	65	66	67	0	0	0	0	72	0	0	0	76	77	78	79	0	81	0	0	0	0	0	0	0	0	0
R2	0	66	67	68	0	0	0	72	0	0	0	76	77	78	79	0	0	82	0	0	0	0	0	0	0	0
R3	65	66	67	0	0	0	0	72	0	0	0	0	0	78	79	80	0	82	0	0	0	0	0	0	0	0
R4	0	66	67	0	0	0	0	72	0	0	0	0	77	78	0	0	0	82	0	0	0	0	0	0	0	0
R5	65	66	0	68	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R6	65	66	67	68	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R7	0	66	0	0	0	0	0	0	0	0	0	0	77	78	79	0	0	0	0	0	0	0	0	0	0	0
R8	0	66	0	0	0	0	0	0	0	0	0	0	0	78	79	80	0	82	83	0	0	0	0	0	0	0
R9	0	66	0	0	0	0	0	72	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R10	0	66	0	0	0	0	0	0	0	74	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R11	65	66	67	0	0	0	0	72	0	74	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R12	65	66	0	68	0	0	0	0	0	74	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R13	0	66	0	0	0	0	0	72	0	0	0	76	77	78	0	0	0	0	0	0	0	0	0	0	0	0
R14	0	66	0	0	0	0	0	0	0	74	0	0	77	78	79	80	0	0	0	0	0	0	0	0	0	0
R15	65	66	67	0	0	0	0	72	0	74	0	0	77	78	79	80	0	0	0	0	0	0	0	0	0	0
R16	65	66	0	68	0	0	0	0	0	74	0	76	77	78	0	0	0	0	0	0	0	0	0	0	0	0
R17	0	0	0	0	0	0	0	72	0	74	0	0	77	78	79	80	0	0	0	0	0	0	0	0	0	0
R18	65	66	67	0	0	0	0	0	0	0	0	0	77	78	0	80	81	82	0	0	0	0	0	0	0	0
R19	0	0	0	68	0	0	0	0	0	74	0	76	77	78	0	0	0	0	0	0	0	0	0	0	0	0
R20	65	0	67	68	0	0	71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

RULES	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
R1	A	B	C					H				L	M	N	O		Q									
R2		B	C	D				H				L	M	N	O			R								
R3	A	B	C					H						N	O	P		R								
R4		B	C					H					M	N				R								
R5	A	B		D																						
R6	A	B	C	D																						
R7		B											M	N	O											
R8		B												N	O	P		R	S							
R9		B						H																		
R10		B								J																
R11	A	B	C					H		J																
R12	A	B		D						J																
R13		B						H				L	M	N												
R14		B								J			M	N	O	P										
R15	A	B	C					H		J			M	N	O	P										
R16	A	B		D						J		L	M	N												
R17								H		J			M	N	O	P										
R18	A	B	C										M	N		P	Q	R								
R19				D						J		L	M	N												
R20	A		C	D			G																			

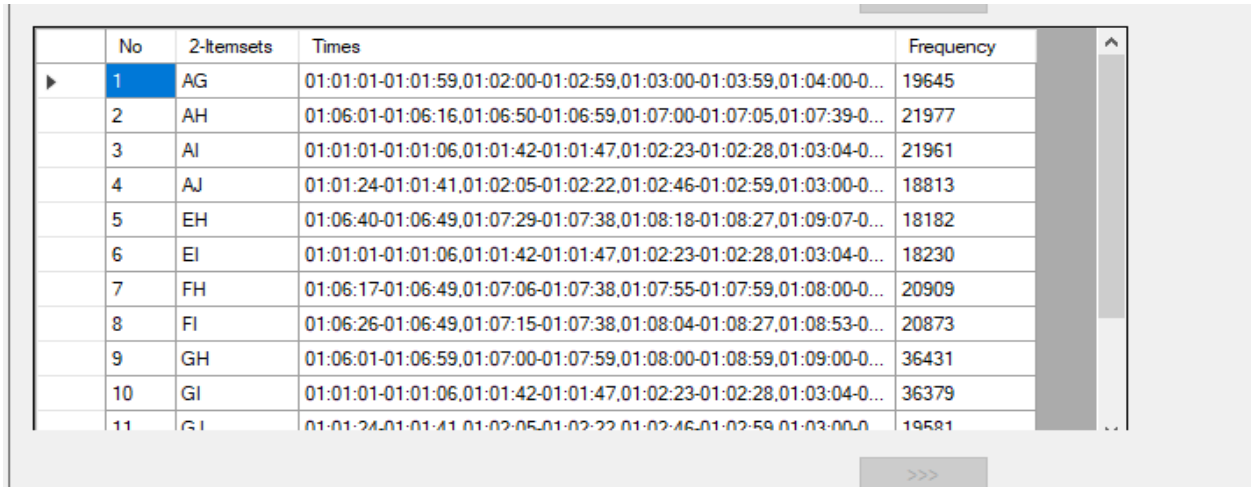
RULES	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
R1	65	66	67	0	0	0	0	72	0	0	0	76	77	78	79	0	81	0	0	0	0	0	0	0	0	0
R2	0	66	67	68	0	0	0	72	0	0	0	76	77	78	79	0	0	82	0	0	0	0	0	0	0	0
R3	65	66	67	0	0	0	0	72	0	0	0	0	0	78	79	80	0	82	0	0	0	0	0	0	0	0
R4	0	66	67	0	0	0	0	72	0	0	0	0	77	78	0	0	0	82	0	0	0	0	0	0	0	0
R5	65	66	0	68	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R6	65	66	67	68	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R7	0	66	0	0	0	0	0	0	0	0	0	0	77	78	79	0	0	0	0	0	0	0	0	0	0	0
R8	0	66	0	0	0	0	0	0	0	0	0	0	0	78	79	80	0	82	83	0	0	0	0	0	0	0
R9	0	66	0	0	0	0	0	72	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R10	0	66	0	0	0	0	0	0	0	74	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R11	65	66	67	0	0	0	0	72	0	74	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R12	65	66	0	68	0	0	0	0	0	74	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R13	0	66	0	0	0	0	0	72	0	0	0	76	77	78	0	0	0	0	0	0	0	0	0	0	0	0
R14	0	66	0	0	0	0	0	0	0	74	0	0	77	78	79	80	0	0	0	0	0	0	0	0	0	0
R15	65	66	67	0	0	0	0	72	0	74	0	0	77	78	79	80	0	0	0	0	0	0	0	0	0	0
R16	65	66	0	68	0	0	0	0	0	74	0	76	77	78	0	0	0	0	0	0	0	0	0	0	0	0
R17	0	0	0	0	0	0	0	72	0	74	0	0	77	78	79	80	0	0	0	0	0	0	0	0	0	0
R18	65	66	67	0	0	0	0	0	0	0	0	0	77	78	0	80	81	82	0	0	0	0	0	0	0	0
R19	0	0	0	68	0	0	0	0	0	74	0	76	77	78	0	0	0	0	0	0	0	0	0	0	0	0
R20	65	0	67	68	0	0	71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

n	T1	xn+1	yn+1	T2
1	65	126.2893	130	0
2	66	124.9499	108.3273	79
3	67	45.33105	28.95351	0
4	0	136.592	18.92294	0
5	0	41.226	50.14894	0
....	....	.....	.....	....
519	0	117.586	135.356	0
520	0	119.1244	107.8579	0

RULES	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
R1	A	B	C					H				L	M	N	O		Q										
R2		B	C	D				H				L	M	N	O			R									
R3	A	B	C					H						N	O	P		R									
R4		B	C					H					M	N				R									
R5	A	B		D																							
R6	A	B	C	D																							
R7		B											M	N	O												
R8		B												N	O	P		R	S								
R9		B						H																			
R10		B								J																	
R11	A	B	C					H		J																	
R12	A	B		D						J																	
R13		B						H				L	M	N													
R14		B								J			M	N	O	P											
R15	A	B	C					H		J			M	N	O	P											
R16	A	B		D						J		L	M	N													
R17								H		J			M	N	O	P											
R18	A	B	C										M	N		P	Q	R									
R19				D						J		L	M	N													
R20	A		C	D			G																				



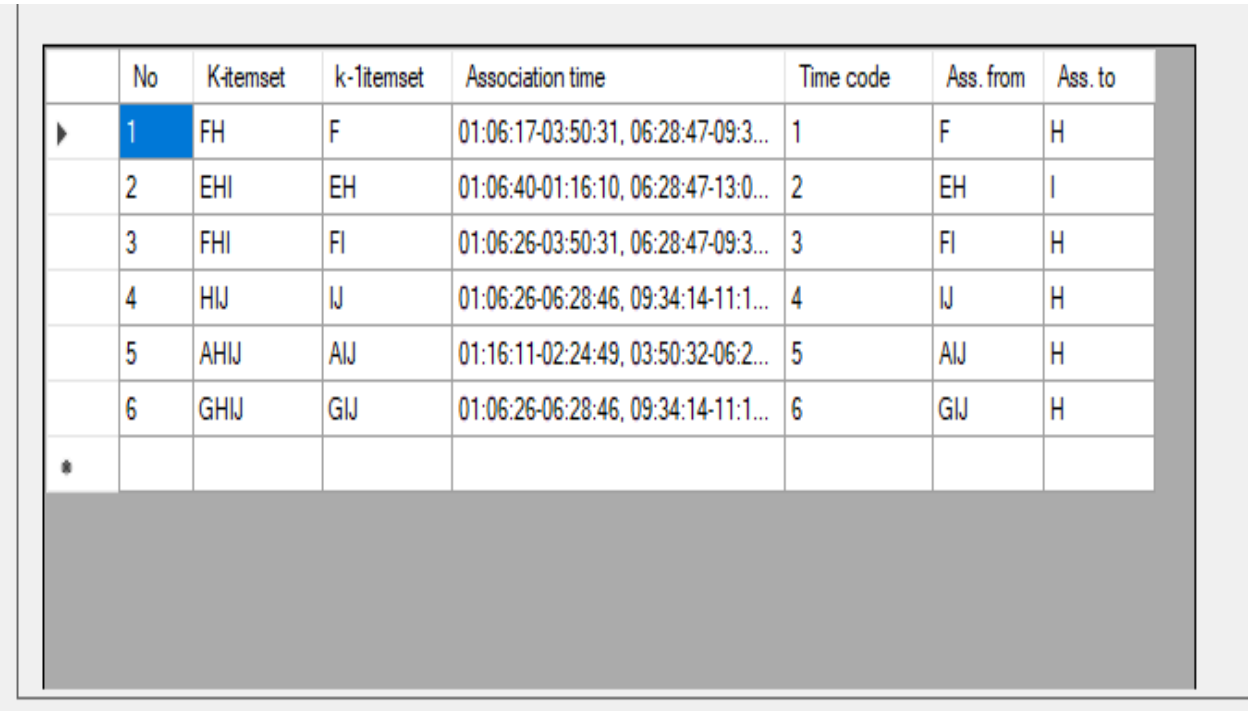
	No	1-Itemsets	Times	Frequency
▶	1	A	01:01:01-01:01:59,01:02:00-01:02:59,01:03:00-01:03:59,01:04:00-0...	22277
	2	E	01:01:01-01:01:23,01:01:42-01:01:59,01:02:00-01:02:04,01:02:23-0...	18356
	3	F	01:06:17-01:06:49,01:07:06-01:07:38,01:07:55-01:07:59,01:08:00-0...	20909
	4	G	01:01:01-01:01:59,01:02:00-01:02:59,01:03:00-01:03:59,01:04:00-0...	36731
	5	H	01:06:01-01:06:59,01:07:00-01:07:59,01:08:00-01:08:59,01:09:00-0...	42886
	6	I	01:01:01-01:01:06,01:01:42-01:01:47,01:02:23-01:02:28,01:03:04-0...	42834
	7	J	01:01:24-01:01:41,01:02:05-01:02:22,01:02:46-01:02:59,01:03:00-0...	24830
*				



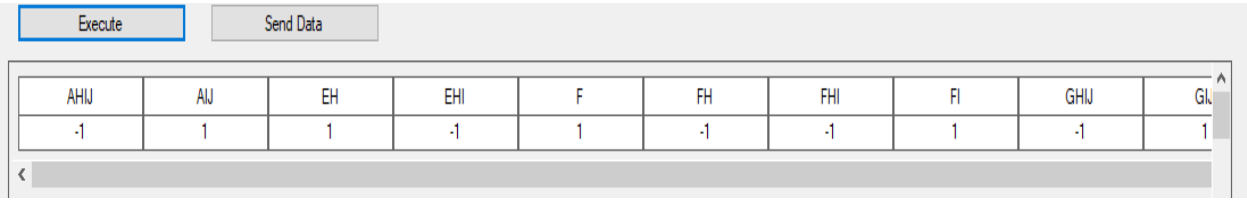
	No	2-Itemsets	Times	Frequency
▶	1	AG	01:01:01-01:01:59,01:02:00-01:02:59,01:03:00-01:03:59,01:04:00-0...	19645
	2	AH	01:06:01-01:06:16,01:06:50-01:06:59,01:07:00-01:07:05,01:07:39-0...	21977
	3	AI	01:01:01-01:01:06,01:01:42-01:01:47,01:02:23-01:02:28,01:03:04-0...	21961
	4	AJ	01:01:24-01:01:41,01:02:05-01:02:22,01:02:46-01:02:59,01:03:00-0...	18813
	5	EH	01:06:40-01:06:49,01:07:29-01:07:38,01:08:18-01:08:27,01:09:07-0...	18182
	6	EI	01:01:01-01:01:06,01:01:42-01:01:47,01:02:23-01:02:28,01:03:04-0...	18230
	7	FH	01:06:17-01:06:49,01:07:06-01:07:38,01:07:55-01:07:59,01:08:00-0...	20909
	8	FI	01:06:26-01:06:49,01:07:15-01:07:38,01:08:04-01:08:27,01:08:53-0...	20873
	9	GH	01:06:01-01:06:59,01:07:00-01:07:59,01:08:00-01:08:59,01:09:00-0...	36431
	10	GI	01:01:01-01:01:06,01:01:42-01:01:47,01:02:23-01:02:28,01:03:04-0...	36379
	11	GJ	01:01:24-01:01:41,01:02:05-01:02:22,01:02:46-01:02:59,01:03:00-0...	19581

No	3-Itemsets	Times	Frequency
1	AGH	01:06:01-01:06:16,01:06:50-01:06:59,01:07:00-01:07:05,01:07:39-0...	19345
2	AGI	01:01:01-01:01:06,01:01:42-01:01:47,01:02:23-01:02:28,01:03:04-0...	19329
3	AHI	01:16:11-01:16:59,01:17:00-01:17:23,01:17:25-01:17:28,01:17:30-0...	21913
4	AHJ	01:06:01-01:06:16,01:06:50-01:06:59,01:07:00-01:07:05,01:07:39-0...	18687
5	AIJ	01:16:11-01:16:59,01:17:00-01:17:23,01:17:25-01:17:28,01:17:30-0...	18623
6	EHI	01:06:40-01:06:49,01:07:29-01:07:38,01:08:18-01:08:27,01:09:07-0...	18182
7	FHI	01:06:26-01:06:49,01:07:15-01:07:38,01:08:04-01:08:27,01:08:53-0...	20873
8	GHI	01:06:26-01:06:49,01:07:15-01:07:38,01:08:04-01:08:27,01:08:53-0...	36331
9	GHJ	01:06:01-01:06:39,01:06:50-01:06:59,01:07:00-01:07:28,01:07:39-0...	19455
10	GIJ	01:06:26-01:06:39,01:07:15-01:07:28,01:08:04-01:08:17,01:08:53-0...	19355
11	HIJ	01:06:26-01:06:39,01:07:15-01:07:28,01:08:04-01:08:17,01:08:53-0	24604

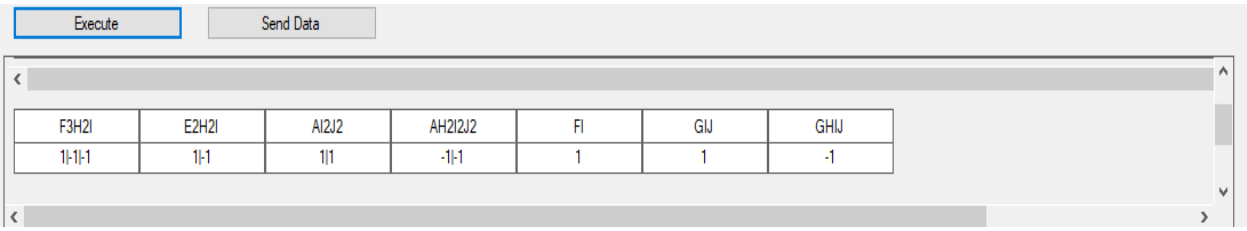
No	4-Itemsets	Times	Frequency
1	AGHI	01:16:33,01:16:40,01:16:47,01:16:54,01:17:01,01:17:08,01:17:15,0...	19281
2	AHIJ	01:16:11-01:16:59,01:17:00-01:17:23,01:17:25-01:17:28,01:17:30-0...	18623
3	GHIJ	01:06:26-01:06:39,01:07:15-01:07:28,01:08:04-01:08:17,01:08:53-0...	19355



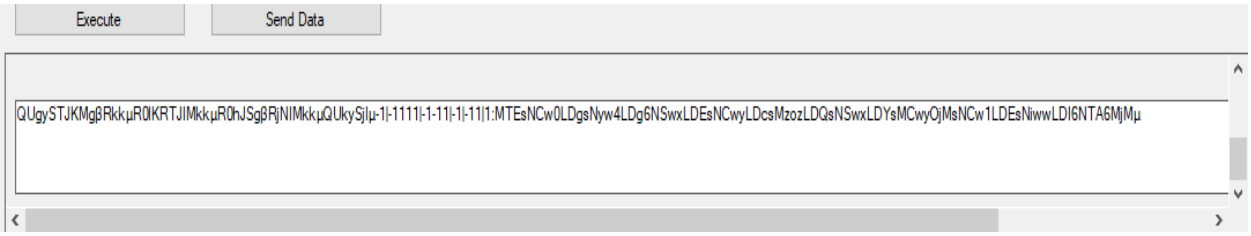
No	K-itemset	k-1itemset	Association time	Time code	Ass. from	Ass. to
1	FH	F	01:06:17-03:50:31, 06:28:47-09:3...	1	F	H
2	EHI	EH	01:06:40-01:16:10, 06:28:47-13:0...	2	EH	I
3	FHI	FI	01:06:26-03:50:31, 06:28:47-09:3...	3	FI	H
4	HIJ	IJ	01:06:26-06:28:46, 09:34:14-11:1...	4	IJ	H
5	AHIJ	AIJ	01:16:11-02:24:49, 03:50:32-06:2...	5	AIJ	H
6	GHIJ	GIJ	01:06:26-06:28:46, 09:34:14-11:1...	6	GIJ	H



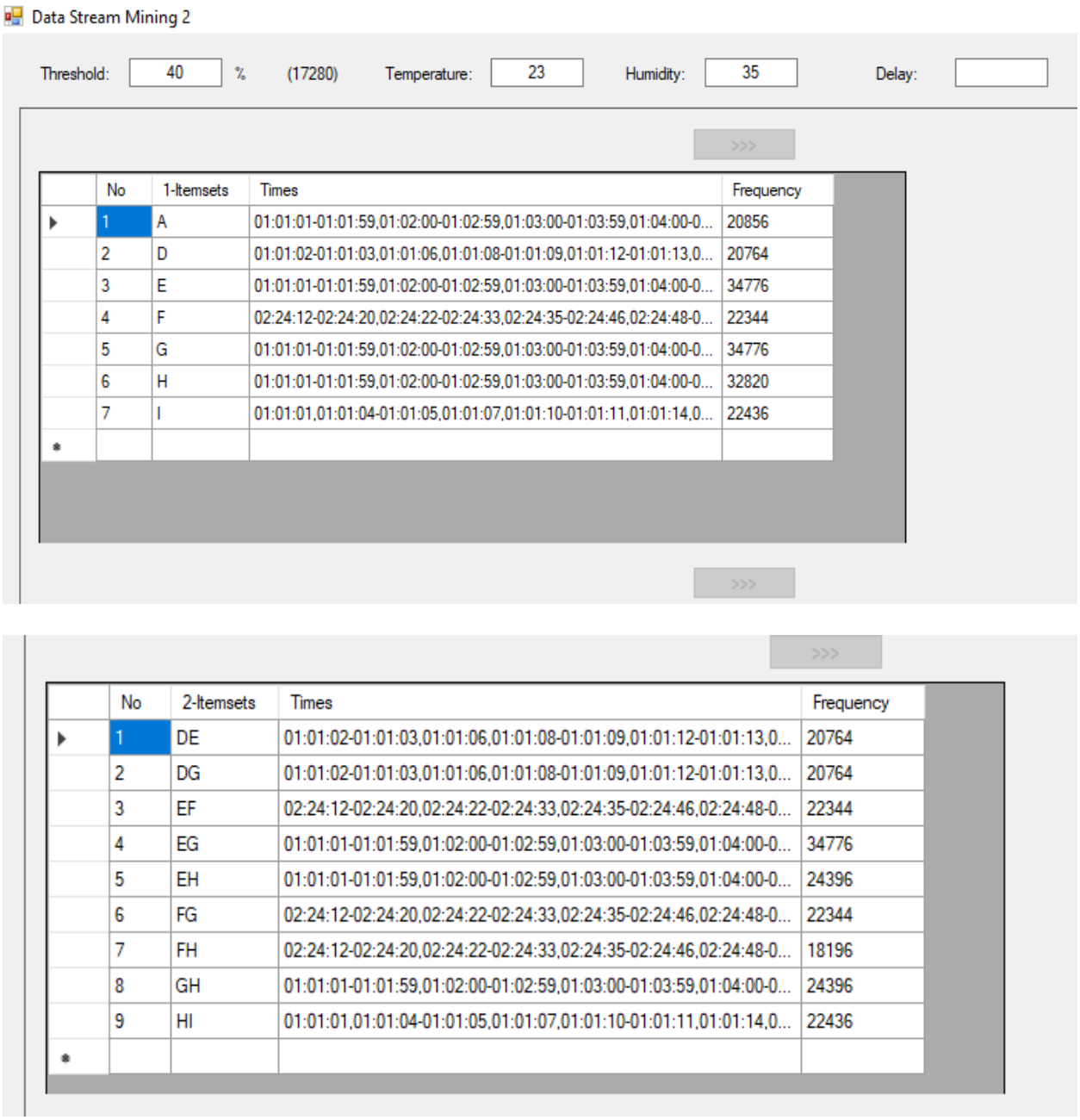
AHIJ	AIJ	EH	EHI	F	FH	FHI	FI	GHIJ	GI
-1	1	1	-1	1	-1	-1	1	-1	1



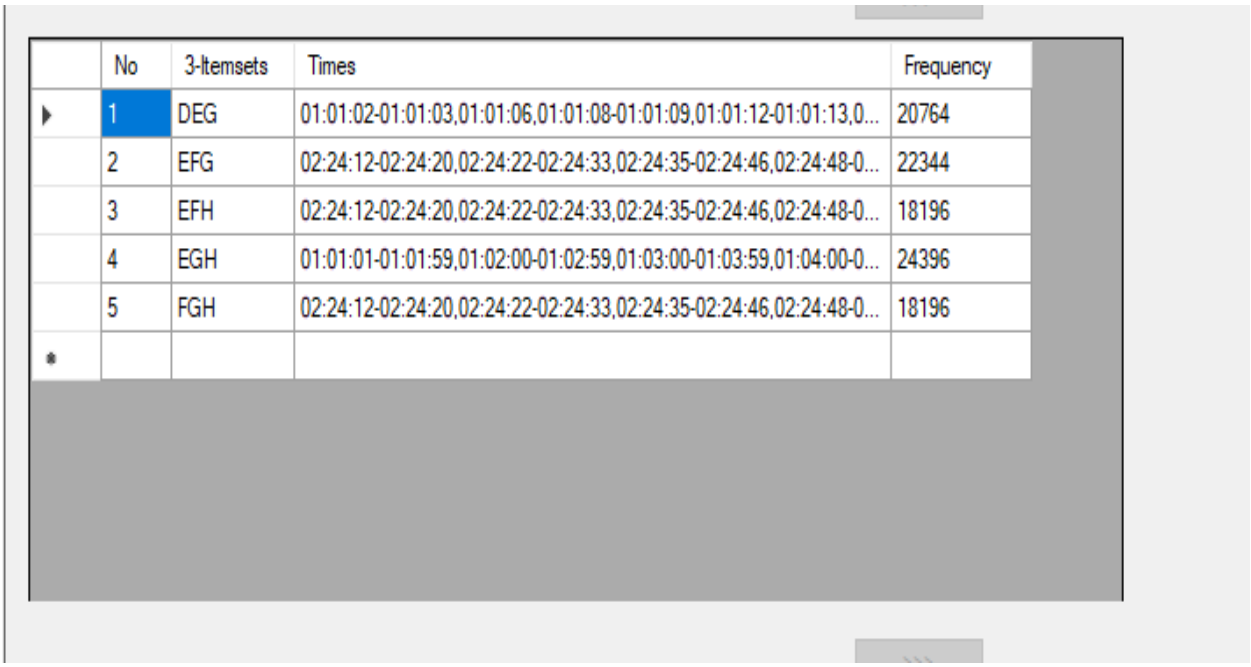
F3H2I	E2H2I	AI2J2	AH2I2J2	FI	GIJ	GHIJ
1|-1|-1	1|-1	1|1	-1|-1	1	1	-1



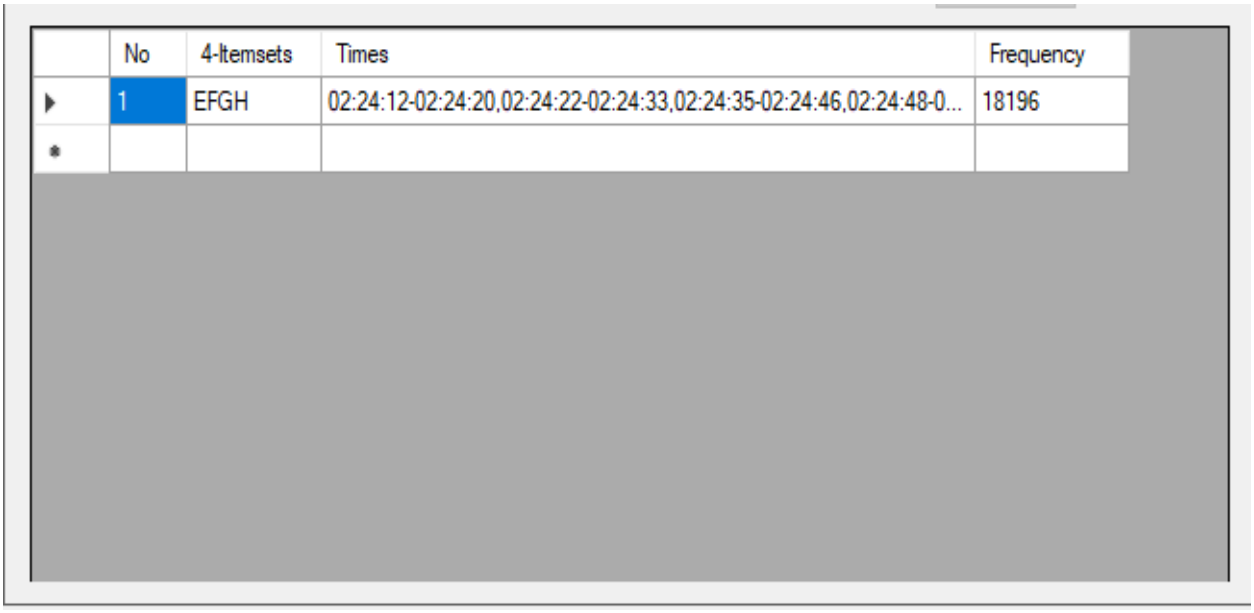
QUgySTJKMgβRkkμR0lKRTJIMkkμR0hJSgβRjNIMkkμQUkySjIμ-1|-1111|-1-11|-1|-11|1:MTEsNCw0LDgsNyw4LDg6NSwxLDEsNCwyLDcsMzozLDQsNSwxLDYsMCwyOjMsNCw1LDEsNiwwLDI6NTA6MjMμ


Data Stream Mining 2
Threshold: 40 % (17280) Temperature: 23 Humidity: 35 Delay:
>>>
No 1-Itemsets Times Frequency
1 A 01:01:01-01:01:59,01:02:00-01:02:59,01:03:00-01:03:59,01:04:00-0... 20856
2 D 01:01:02-01:01:03,01:01:06,01:01:08-01:01:09,01:01:12-01:01:13,0... 20764
3 E 01:01:01-01:01:59,01:02:00-01:02:59,01:03:00-01:03:59,01:04:00-0... 34776
4 F 02:24:12-02:24:20,02:24:22-02:24:33,02:24:35-02:24:46,02:24:48-0... 22344
5 G 01:01:01-01:01:59,01:02:00-01:02:59,01:03:00-01:03:59,01:04:00-0... 34776
6 H 01:01:01-01:01:59,01:02:00-01:02:59,01:03:00-01:03:59,01:04:00-0... 32820
7 I 01:01:01,01:01:04-01:01:05,01:01:07,01:01:10-01:01:11,01:01:14,0... 22436
>>>
No 2-Itemsets Times Frequency
1 DE 01:01:02-01:01:03,01:01:06,01:01:08-01:01:09,01:01:12-01:01:13,0... 20764
2 DG 01:01:02-01:01:03,01:01:06,01:01:08-01:01:09,01:01:12-01:01:13,0... 20764
3 EF 02:24:12-02:24:20,02:24:22-02:24:33,02:24:35-02:24:46,02:24:48-0... 22344
4 EG 01:01:01-01:01:59,01:02:00-01:02:59,01:03:00-01:03:59,01:04:00-0... 34776
5 EH 01:01:01-01:01:59,01:02:00-01:02:59,01:03:00-01:03:59,01:04:00-0... 24396
6 FG 02:24:12-02:24:20,02:24:22-02:24:33,02:24:35-02:24:46,02:24:48-0... 22344
7 FH 02:24:12-02:24:20,02:24:22-02:24:33,02:24:35-02:24:46,02:24:48-0... 18196
8 GH 01:01:01-01:01:59,01:02:00-01:02:59,01:03:00-01:03:59,01:04:00-0... 24396
9 HI 01:01:01,01:01:04-01:01:05,01:01:07,01:01:10-01:01:11,01:01:14,0... 22436

O



No	3-Itemsets	Times	Frequency
1	DEG	01:01:02-01:01:03,01:01:06,01:01:08-01:01:09,01:01:12-01:01:13,0...	20764
2	EFG	02:24:12-02:24:20,02:24:22-02:24:33,02:24:35-02:24:46,02:24:48-0...	22344
3	EFH	02:24:12-02:24:20,02:24:22-02:24:33,02:24:35-02:24:46,02:24:48-0...	18196
4	EGH	01:01:01-01:01:59,01:02:00-01:02:59,01:03:00-01:03:59,01:04:00-0...	24396
5	FGH	02:24:12-02:24:20,02:24:22-02:24:33,02:24:35-02:24:46,02:24:48-0...	18196



No	4-Itemsets	Times	Frequency
1	EFGH	02:24:12-02:24:20,02:24:22-02:24:33,02:24:35-02:24:46,02:24:48-0...	18196

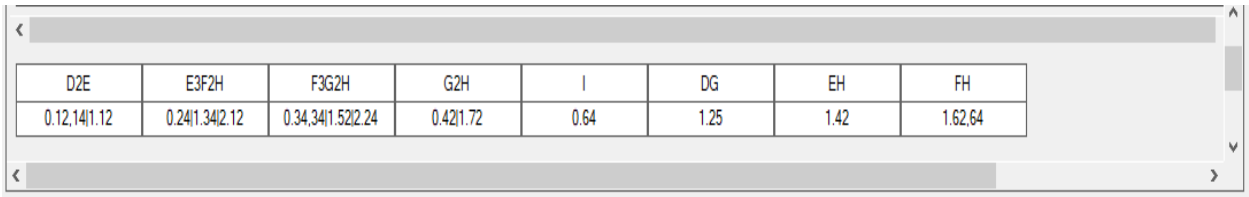
No	K-itemset	k-1itemset	Association time	Time code	Ass. from	Ass. to
1	DE	D	01:01:02-01:34:32, 02:24:13-06:5...	1	D	E
2	DG	D	01:01:02-01:34:32, 02:24:13-06:5...	1	D	G
3	EF	F	02:24:12-02:57:03, 03:49:00-06:2...	2	F	E
4	HI	I	01:01:01-02:57:02, 04:23:36-08:1...	3	I	H
5	FG	F	02:24:12-02:57:03, 03:49:00-06:2...	2	F	G
6	EG	E	01:01:01-01:34:32, 02:24:12-06:5...	4	E	G
7	EG	G	01:01:01-01:34:32, 02:24:12-06:5...	4	G	E
8	EFH	FH	02:24:12-02:57:03, 04:23:34-06:2...	5	FH	E
9	FGH	FH	02:24:12-02:57:03, 04:23:34-06:2...	5	FH	G
10	DEG	DE	01:01:02-01:34:32, 02:24:13-06:5...	1	DE	G
11	DEG	DG	01:01:02-01:34:32, 02:24:13-06:5	1	DG	E

No	K-itemset	k-1itemset	Association time	Time code	Ass. from	Ass. to
9	FGH	FH	02:24:12-02:57:03, 04:23:34-06:2...	5	FH	G
10	DEG	DE	01:01:02-01:34:32, 02:24:13-06:5...	1	DE	G
11	DEG	DG	01:01:02-01:34:32, 02:24:13-06:5...	1	DG	E
12	EGH	EH	01:01:01-01:34:32, 02:24:12-02:5...	6	EH	G
13	EGH	GH	01:01:01-01:34:32, 02:24:12-02:5...	6	GH	E
14	EFG	EF	02:24:12-02:57:03, 03:49:00-06:2...	2	EF	G
15	EFG	FG	02:24:12-02:57:03, 03:49:00-06:2...	2	FG	E
16	EFGH	FGH	02:24:12-02:57:03, 04:23:34-06:2...	5	FGH	E
17	EFGH	EFH	02:24:12-02:57:03, 04:23:34-06:2...	5	EFH	G

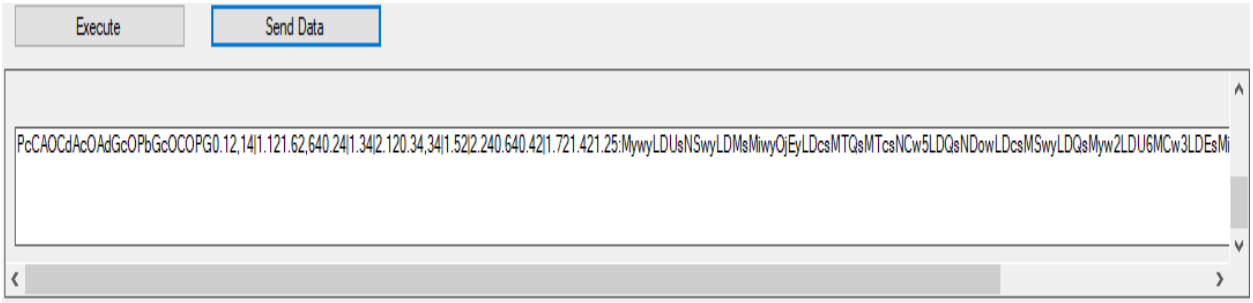
Execute Send Data

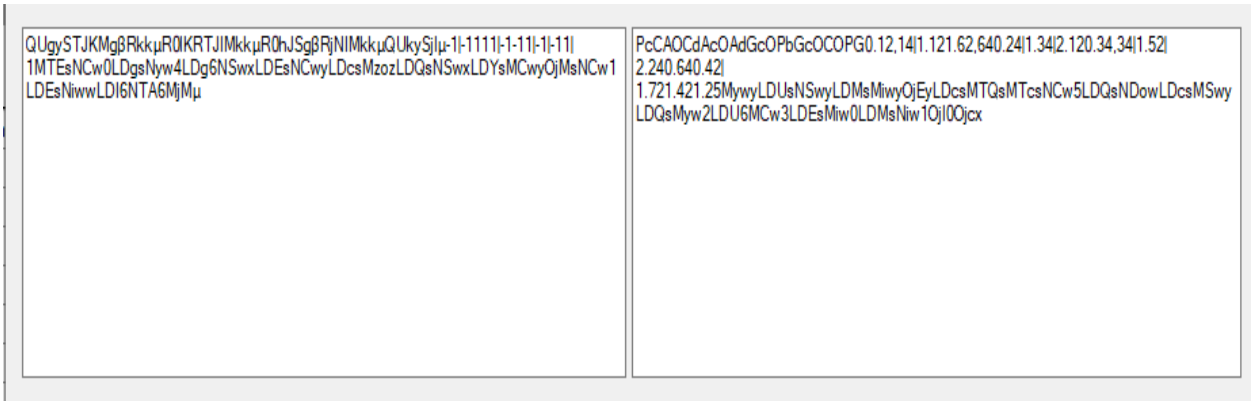
D	DE	DG	E	EF	EFH	EH	F	FG	FGH
0.12,14	1.12	1.25	0.24	1.34	2.12	1.42	0.34,34	1.52	2.2

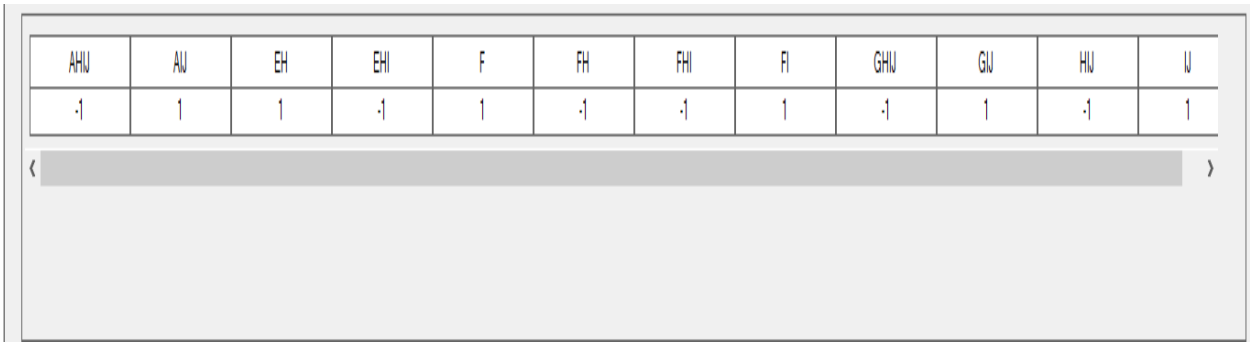
Q



D2E	E3F2H	F3G2H	G2H	I	DG	EH	FH
0.12,14|1.12	0.24|1.34|2.12	0.34,34|1.52|2.24	0.42|1.72	0.64	1.25	1.42	1.62,64







AHJ	AJ	EH	EHI	F	FH	FHI	FI	GHJ	GJ	HJ	IJ
-1	1	1	-1	1	-1	-1	1	-1	1	-1	1

R

D2E	E3F2H	F3G2H	G2H	I	DG	EH	FH
0.12,14|1.12	0.24|1.34|2.12	0.34,34|1.52|2.24	0.42|1.72	0.64	1.25	1.42	1.62,64

D	DE	DEG	DG	E	EF	EFG	EFGH	EFH	EG	EGH	EH
2	11	-2	11	1	11	-2	-2	11	-2	-2	1

AHJ	AJ	D	DE	DEG	DG	E	EF	EFG	EFGH	EFH	EG
-1	1	2	11	-2	11	1	11	-2	-2	11	-2

	No	1-itemset	2-itemset
▶	1	D	DE
	2	D	DG
	3	E	EF
	4	E	EG
	5	F	FG
	6	F	FH
	7	G	EG
	8	I	HI
*			

S

No	2-itemset	3-itemset
1	DE	DEG
2	DG	DEG
3	EF	EFG
4	EF	EFH
5	EH	EGH
6	EH	EHI
7	FG	EFG
8	FG	FGH
9	FH	FHI

No	3-itemset	4-itemset
1	AIJ	AHIJ
2	EFH	EFGH
3	FGH	EFGH
4	GIJ	GHIJ

T

No	K-itemset	K+1itemset	Ass. from	Ass. to
1	D	DE	D	E
2	D	DG	D	G
3	E	EF	E	F
4	E	EG	E	G
5	F	FG	F	G
6	F	FH	F	H
7	G	EG	G	E
8	I	HI	I	H
9	DE	DEG	DE	G
10	DG	DEG	DG	E
11	EF	EFG	EF	G
11	EF	EFG	EF	G
12	EF	EFH	EF	H
13	EH	EGH	EH	G
14	EH	EHI	EH	I
15	FG	EFG	FG	E
16	FG	FGH	FG	H
17	FH	FHI	FH	I
18	GH	EGH	GH	E
19	IJ	HIJ	IJ	H
20	AIJ	AHIJ	AIJ	H
21	EFH	EFGH	EFH	G
22	FGH	EFGH	FGH	E
23	GIJ	GHIJ	GIJ	H

U

الملخص

مع التطور السريع لتقنيات المعلومات وتعدين البيانات ، يميل العديد من الباحثين إلى حل مشكلة مركزية تتعلق بتعدين البيانات ، وهي حفظ الخصوصيه لتعدين البيانات. يعتبر تعدين البيانات في البيئات الذكية التي تعتمد على المستشعرات أحد أهم وأحدث أنواع التنقيب عن البيانات في الآونة الأخيرة.

تعدين قواعد الارتباط يحدد ما هو مخفي عن العلاقات الترابطيه بين العناصر والسمات في اي بيئه. تعتبر هذه المعلومات مهمه و حساسة ، والكشف عن هذه المعرفة الهامة للعموم او المنافسين تعتبر مشكلة مركزية.

تعتبر تقنية حفظ خصوصيه تعدين البيانات تقنية مهمة وحديثة ، خاصة للعديد من التطبيقات. حفظ خصوصيه تعدين البيانات هو عبارة عن مزيج من تخصصين مهمين: اكتشاف المعرفة وأمن البيانات.

تمت دراسة حفظ خصوصيه تعدين البيانات على نطاق واسع مع البيئات المركزية غير الموزعة ، خاصة مع البيانات الثابتة ، حيث لا يوجد عامل الوقت وحتى في حالة وجود عامل الوقت ضمن مجموعة البيانات ، فسيتم إهماله أثناء خطوات ونتائج التعدين.

أما بالنسبة لنظامنا المقترح ، فهو مصمم للتعامل مع تدفقات البيانات من أجهزة الاستشعار في البيئات الذكية الموزعة ، حيث يتم أخذ عامل الوقت في الاعتبار وإدراجه في الخطوات الرئيسية لخوارزمية التعدين.

يتضمن النظام المقترح خوارزمية تعدين البيانات التي تعمل في كل المواقع في نفس الوقت. بعد الحصول على نتائج التعدين (قواعد الارتباط مصحوبة بأوقات الارتباط) ، يتم استخدام خوارزميتين لضغط وترميز نتائج التعدين في المواقع الموزعة المختلفة (خوارزمية في كل موقع).

أظهرت النتائج التجريبية أن خوارزمية التعدين تقلل الوقت المطلوب بمقدار 4/1 من الوقت المطلوب في حالة استخدام خوارزمية التعدين القياسية (خوارزمية Apriori)، كما تقلل من مساحة التخزين المطلوبة ايظا. بالاضافه الى ذلك والاهم ظهر في هذه الاطروحه مصطلح جديد لم يكن يعرف سابقا وهو اوقات الارتباط الذي يحول نتائج التعدين للبيانات المتدفقه من التخمينيه الى المظبوطه ومحدده باوقات معينه.

أما بالنسبة للنتائج النهائية في الموقع المسؤول عن حساب قواعد الارتباط لجميع المواقع ، فقد تم الحصول على قواعد الارتباط بالكامل (100%) دون فقدان أي من قواعد الارتباط أو ظهور قواعد ارتباط مزيفة.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل- كلية تكنولوجيا المعلومات

تحسين خصوصية قواعد الارتباط بالاعتماد على تقنيات الضغط والترميز

اطروحة مقدمة
الى مجلس كلية تكنولوجيا المعلومات- جامعة بابل وهي جزء من متطلبات نيل درجة الدكتوراة فلسفة في تكنولوجيا المعلومات / برمجيات

من قبل

وحيد عبد الكاظم سلمان

باشراف

الأستاذ الدكتور المهندس ستار بدر سدخان

2021 A.D. **1443 A.H.**