*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 11s*
*DOI: https://doi.org/10.17762/ijritcc.v11i11s.8175*
***Article Received:** 30 June 2023 **Revised:** 24 August 2023 **Accepted:** 06 September 2023*

# Clustering of Non-Associated Item Sets for Analyzing Show Room Sales Dataset

**Vinayababu M[1], M Sreedevi[2]**
[1]Research Scholar, Department of Computer Science,
Sri Venkateswara University, Tirupati, India;
vinayababu955@gmail.com
[2]Department of Computer Science,
Sri Venkateswara University, Tirupati, India;
msreedevi_svu2007@yahoo.com

**Abstract**— Market basket analysis (MBA) is a well-liked method for identifying relationships between products that people purchase in a database. It is predicated on association rule mining (ARM), a data mining technique that pulls valuable data from huge databases. Due to consumers using internet applications for online shopping and insurance, an ever-increasing amount of data is generated online. It produces large amounts and, if mined effectively, will greatly benefit society as a whole as well as individuals. So, numerous data science and machine learning-related techniques have been created to gradually unlock the potential. The Clustering of Non-Associated Item Sets (CNAIS) of the Sales dataset used in the Showroom for choosing customers for benefits and web application design is discussed in this study. The CNAIS algorithm implementation process and dataset for this study are discussed.

**Keywords**- Item Sets, Clustering, Show Room Sales, machine learning, CNAIS algorithm.

## I. INTRODUCTION

An effective tool for making data-based judgements is association rules. They can find common and significant patterns in a variety of fields, including business, medicine, the web, and DNA analysis. For instance, by determining which products are frequently purchased together, association rules can assist businesses in strengthening their marketing plans [1]. By identifying common symptoms or risk factors, they can also assist medical practitioners in the diagnosis of diseases. Association rules can also be utilised for gene expression analysis in DNA data, fraud detection in web transactions, and condition monitoring of high voltage equipment in electrical power engineering. Many diverse domains of knowledge can benefit from and apply association rules. Analysis of web users' navigational activity based on their log data is one of the uses of association rule mining. Several techniques, including Apriori, ECLAT, Tree Projection, FP-growth, direct hashing, and pruning, can be used to find association rules from data. The frameworks for routine pattern mining and the computational effectiveness of these methods vary.

Market basket analysis (MBA) is a well-liked method for identifying relationships between products that people purchase in a database. It is predicated on association rule mining (ARM), a data mining technique that pulls valuable data from huge databases. The objective of ARM is to quickly locate the most intriguing and pertinent patterns [2]. By examining the transactions that involve one or more items, MBA can assist retailers in better understanding customer behaviour and preferences. For instance, there is a higher likelihood that a customer will purchase butter if he purchases bread. This form of association rule can be applied to recommend product pairings, create better retail layouts, and boost co-branding and brand loyalty. Additionally, it can assist managers in developing practical and effective plans. MBA scans bar codes to record information about each goods purchased. Given that it uncovers the hidden patterns of consumer spending, this data has enormous potential value for marketing [3].

The internet is frequently used to automatically extract knowledge from web log files. This can assist information providers in better comprehending the wants and needs of web users and in optimizing websites by customizing their structure and content to match user behaviours [3]. Finding the appropriate tools to examine online log data and extract important information is difficult, though. Many of the current web log analysis solutions lack user-friendliness and are viewed as being slow, rigid, expensive, difficult to maintain, or having limited findings. Applying data mining techniques, sometimes referred to as web log mining or web usage mining, is a novel alternative for identifying process trends in web data. When compared to conventional decision-making methods, these methodologies can provide additional insights. mining patterns frequently

### 2.1 Types of Clustering Techniques

Various algorithms are processed in the evaluation, including their adaptability to different data sets, capacity to handle high dimensionality, sophistication, and applicability in multiple

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 11s*
*DOI: https://doi.org/10.17762/ijritcc.v11i11s.8175*
***Article Received:** 30 June 2023 **Revised:** 24 August 2023 **Accepted:** 06 September 2023*

fields. Accordingly, the based on mechanism of processing datasets clustering techniques could be widely categorized as follows:

**a) Partitioning-based:** These clustering algorithms are simple to implement, and clusters are easily determined effectively. Initially, groups are specified and finally reallocated and combined to form clusters. In other words, algorithms like K-Means, K-Medoids, P.A.M., CLARA, CLARANS and FCM scan and divide data objects into several partitions where each partition is considered a cluster[4].

**b) Hierarchical-based:** Also known as Agglomerative or divisive partitioning. By considering proximities in attributes of the data set, clusters are hierarchically organized either top-to-down. The intermediate nodes procure proximities. A dendogram is formed as the algorithm scans the datasets, with leaf nodes representing individual data. The initial cluster gradually expands into the hierarchical tree to form several clusters as new nodes are added. The approach in Hierarchical clustering methodologies is either bottom-up or top-down. BIRCH, CURE, ROCK, and Chameleon are some of the algorithms of this technique [5].

**c) Density-based:** Build the clusters regions of density, connectivity and boundary are considered when processing data objects. A cluster is a connected dense component that grows in any direction depending on density. It is shown as a point nearest to neighbors. Clusters of variable shapes are created in these density-based algorithms like DBSCAN, OPTICS, DBCLASD and DENCLUE [6].

**d) Grid-based:** The database on which clustering is to be done divides the space of the data objects into grids. The grid-based approach is efficient in terms of time complexity because it performs a single pass through the dataset to calculate the grid's statistical values. Due to this, grid-based algorithms are independent of dataset size. Examples of this category typically include Wave-Cluster and STING [7].

**e) Model-based:** These robust clustering techniques choose the best fit between the supplied data and a particular (predefined) mathematical model by considering the number of outliers. Statistics and neural networks are the two main methods taken into account in model-based methods. According to the evaluation, the MCLUST technique is likely the most well-known model based on complexity and cluster fitting [8-10].

## II. RELATED WORK

Data mining is the process of finding or removing significant knowledge, information, or patterns from data in huge databases. Data mining techniques can be used in a variety of fields, including medicine, where they can aid in the diagnosis and treatment of disorders [11]. Data regarding healthcare is gathered and analyzed in the healthcare sector to enhance patient outcomes and service quality. Using association rule (AR) based Apriori Algorithm, one of the uses of data mining in healthcare is to find the most prevalent diseases in a specific geographic area and time period.

The clustering technique involves following processes such as
a) Feature extraction and selection: extract and select the most representative features from the original data set;
b) Clustering algorithm design: design the clustering algorithm according to the characteristics of the problem;
c) Evaluation: evaluate the clustering result to verify the validity of the algorithm by using practically available dataset.

### 2.1 Types of Clustering Techniques

Various algorithms are processed in the evaluation, including their adaptability to different data sets, capacity to handle high dimensionality, sophistication, and applicability in multiple fields [12]. Accordingly, the based on mechanism of processing datasets clustering techniques could be widely categorized as follows:

**a) Partitioning-based:** These clustering algorithms are simple to implement, and clusters are easily determined effectively. Initially, groups are specified and finally reallocated and combined to form clusters. In other words, algorithms like K-Means, K-Medoids, P.A.M., CLARA, CLARANS and FCM scan and divide data objects into several partitions where each partition is considered a cluster [13].

**b) Hierarchical-based:** Also known as Agglomerative or divisive partitioning. By considering proximities in attributes of the data set, clusters are hierarchically organized either top-to-down. The intermediate nodes procure proximities. A dendogram is formed as the algorithm scans the datasets, with leaf nodes representing individual data. The initial cluster gradually expands into the hierarchical tree to form several clusters as new nodes are added. The approach in Hierarchical clustering methodologies is either bottom-up or top-down. BIRCH, CURE, ROCK, and Chameleon are some of the algorithms of this technique [14].

**c) Density-based:** Build the clusters regions of density, connectivity and boundary are considered when processing data objects. A cluster is a connected dense component that grows in any direction depending on density. It is shown as a point nearest to neighbors. Clusters of variable shapes are created in these density-based algorithms like DBSCAN, OPTICS, DBCLASD and DENCLUE [15].

**d) Grid-based:** The database on which clustering is to be done divides the space of the data objects into grids. The grid-based approach is efficient in terms of time complexity because it performs a single pass through the dataset to calculate the grid's statistical values. Due to this, grid-based algorithms are independent of dataset size. Examples of this category typically include Wave-Cluster and STING [16].

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 11s*
*DOI: https://doi.org/10.17762/ijritcc.v11i11s.8175*
***Article Received:** 30 June 2023 **Revised:** 24 August 2023 **Accepted:** 06 September 2023*

**e) Model-based:** These robust clustering techniques choose the best fit between the supplied data and a particular (predefined) mathematical model by considering the number of outliers. Statistics and neural networks are the two main methods taken into account in model-based methods. According to the evaluation, the MCLUST technique is likely the most well-known model based on complexity and cluster fitting.

### 2.2 K-Means Clustering Technique

Among various clustering techniques, here we discuss the basic K-means algorithm, a simple, efficient and most used unsupervised clustering algorithm to cluster the data. It is one of the partition-dependent methodologies. The algorithm is performed in two stages.

A) In stage-1 the data is first divided into k clusters, with k's value being predetermined . Take 'k' points from the set of data provided and use them as the centroid for each cluster [17].

b) In Stage-2 determine the distance among the point and centroid, then assign the point to the cluster with the shortest distance, bringing the point's centroid as close as possible. The number of iterations and the shifting of point locations within clusters are decreased by using this method[18].

### 2.3 Cross Industry Standard Process in Data Mining (CRISP-DM)

CRISP-DM is an industry-independent process for data mining. It was developed by Daimler Chrysler (then Daimler-Benz),SPSS (then ISL) and NCR in 1999 which is an industry tool application neutral data mining model build with help of data mining engineers and mining tools service providers. This model allows to find a better and faster results from data mining process[19-20].

### 2.2 Cross Industry Standard Process in Data Mining (CRISP-DM)

CRISP-DM is an industry-independent process for data mining. It was developed by Daimler Chrysler (then Daimler-Benz),SPSS (then ISL) and NCR in 1999 which is an industry tool application neutral data mining model build with help of data mining engineers and mining tools service providers [19-22]. This model allows to find a better and faster results from data mining process. CRISP-DM model is shown in figure 1.

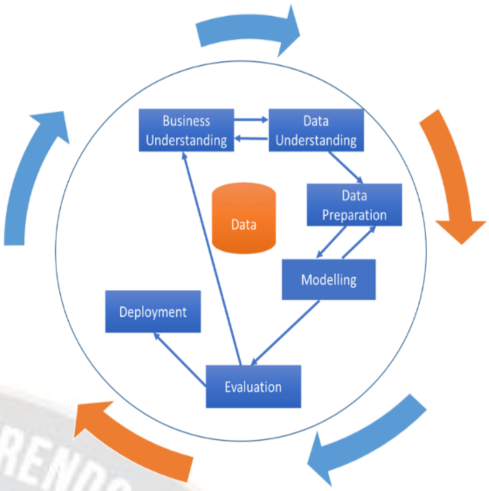


Figure 1. Phases in CRISP-DM

The CRISP-DM model for data mining contains various phases which allows performs steps of the life cycle of a data mining project. It contains the following Six phases such as

a) Business understanding – it lays us the objectives to be satisfied and requirements in view of business perspective. Allows us to know resource availability, assess risk and contingencies, essential output to generate and tools that can be applied to achieve objective.

b) Data Understanding - it used to not only identify, collect, and analyse the data sets but also we can know how data is described /represented to query it, clean the data to remove unnecessary data like missing values, duplicate value which is helpful to improve accuracy in final result.

c) Data Preparation- in this phase data sets are prepared for final data set used for data modelling. It contains five tasks such as selection data (which features are essential), cleaning of data to remove unnecessary values or error values, construction of data such as deriving new columns for computation purpose, integration of data from different resources if required, reformat data if it is essential.

d) Modelling: - Modelling phase allows us to select which algorithms are to be used against data sets to get desired output. Data sets are to be partitioned into training set and test set for validating output from model selected. Model results are to be interpreted based on domain knowledge, the pre-defined success criteria, and the test design.

e) Evaluation: - this phase allows to choose which model is best suitable to meet business needs, to analysis final summary and findings and decide whether results are sufficient or to iterate the process for refinement of

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 11s***
***DOI: https://doi.org/10.17762/ijritcc.v11i11s.8175***
***Article Received:** 30 June 2023 **Revised:** 24 August 2023 **Accepted:** 06 September 2023*

results or to change of model is essential is decided or can go for new project.

f) Deployment: - this Final phase of CRISP-DM will provide deliverable to enterprise such as reports with summaries/findings. Allows to plan how to deploy model, how to maintain and monitor the model. Model is reviewed whether is it right model or whether any changes is essential in future is discussed.

## III. IMPLEMENTATION METHODOLOGY OF CNIAS

The two major goals of data mining are either prediction or description. While description data mining looks for patterns in data to analyse them, prediction data mining forecasts unknown future outcomes using various aspects of the dataset [21-22]. Figure 1 shows that, although it is only one step in the process of uncovering hidden information in data, data mining is the essential step. Data must first be gathered from many sources, integrated, and stored in a data warehouse before any data mining can begin. Data mining algorithms are then used to analyse the data to look for unnoticed patterns once the stored raw data has been assessed and selected using pre-processing processes to provide a standard format.

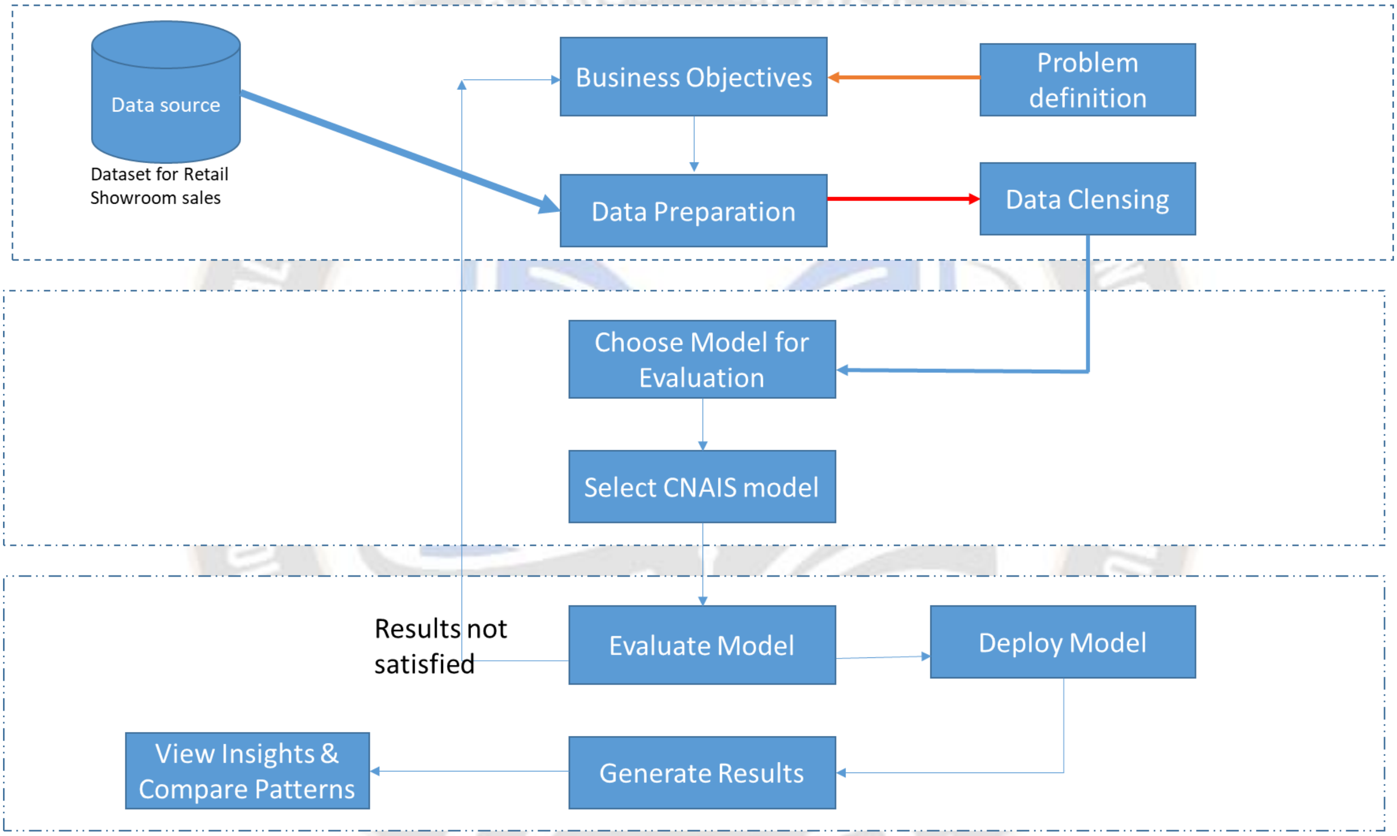


Figure: 1 Functional diagram of Data mining Process based on CRISP-DM

### 3.1 Business Understanding

The first step of the CRISP-DM process is business understanding or domain understanding. By utilising various analytical, processing, and implementation methodologies, it entails determining the business area that needs to be translated into relevant information. Here, the term "problem" is used as The initial document or base for business understanding also includes describing the issues and goals, as well as evaluating the hardware and human resources that are available. The business goals are in line with the goals of the thesis, and figure 1 lists the numerous tasks that need to be completed.

### 3.2Data pre-processing

Every research project or data science-based research endeavor starts with the time-consuming stage of data preparation. On average, data preparation takes up 60% to 70% of the time in a data science project. Data collection and cleansing against missing or null values, among other things, are part of the data preparation process. There are two types of data gathering sources: external sources and internal sources. While internal source data may be simple and inexpensive, it could also contain useless information. Outsourced data may have drawbacks including high cost, long duration, and low quality. The process of gathering data also reveals whether the data is

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 11s***
***DOI: https://doi.org/10.17762/ijritcc.v11i11s.8175***
***Article Received:** 30 June 2023 **Revised:** 24 August 2023 **Accepted:** 06 September 2023*

qualitative (categorical) or quantitative (continuous, count). Additionally, it provides information about the dataset's balance or unbalance. The examination of the data source and its properties occurs in the data description stage of data analysis. Identifying the quantity, structure, and format of the data as well as the type of data source—such as RDBMS, SQL, NoSQL, Big data, etc.—is required. The accuracy and effectiveness of the analysis are impacted by the quantity of the data set. Understanding the data also requires knowledge of the quantity of records, tables, databases, variables, and data kinds (numeric, categorical, or Boolean). The accessibility and availability of the attributes that are important for the analysis are also checked in the data description. In order to improve the quality and usability of the data, data cleaning is a crucial stage in the data mining process. It does this by identifying and eliminating errors, inconsistencies, outliers, and missing information. Depending on the size and complexity of the data set, data cleaning can be done manually or automatically. Data cleansing can increase the trustworthiness and validity of the results as well as the precision and efficiency of data mining algorithms. Data exploration, visualization, analysis, and interpretation can all be aided by data cleansing.

### 3.3Model Selection & Assessment

Accuracy and performance are factors that impact model selection strategies. When using a data mining model selection technique when the output variable is known, supervised learning must be used. The initial method of choice when parameter interpretation is crucial is regression. At this stage, the model is evaluated based on errors and accuracy, and the error and accuracy are tuned in an appropriate way.
Different methods can be used to evaluate the model. The first stage is to assess the model's effectiveness and performance in light of the desired results. It is important to gauge and assess the model's accuracy, which represents its repeatability and reproducibility. The model should also be easily deployable, robust, scalable, and maintainable. The evaluation should confirm that the model fits the criteria and yields satisfactory results (with balanced precision, recall, and sensitivity).

### 3.4 Evaluation

Ranking the models that use the same dataset according to their quality, simplicity, and deployment cost is part of the evaluation process. The examination of the data adequacy, the ideas, comments, and advice from the solutions team and the subject matter experts, as well as the documenting of these in the organizational process assets, are all included in the evaluation report.

### 3.5 Deployment

PEST (political, economic, social, and technological) aspects should be taken into account during the deployment process because they have an impact on the organisation and its surroundings. whereas PEST and SWOT (strengths, weaknesses, opportunities, and threats) analysis are similar, PEST focuses on external factors whereas SWOT analyses internal factors. The deployment procedures ought to guarantee dependability and uniformity from development to production. The hardware, software, and human resource requirements should all be specified in the deployment plan. The deployment strategy should also include maintenance and monitoring techniques to assess the model's validity and performance and to carry out retirement, replacement, or update actions as necessary.

### 3.6 Viewing Insights & Patterns in Results

In order to extract useful information from huge and complex databases, data mining technologies are crucial. They make it possible for analysts and researchers to find unobserved linkages, trends, and patterns that can guide decision-making and problem-solving. Anomalies, outliers, and errors that could normally go unreported can be found using data mining technologies. The following are some advantages of employing data mining tools:

- They are capable of handling enormous amounts of data fast and effectively.
- They have the ability to use a variety of algorithms and strategies to accommodate varied data kinds and goals.
- They have the ability to intuitively and interactively visualise and present the outcomes.
- By automating and streamlining the data analysis process, they may save time and money.

In the data preparation phase, errors in the data must be found and fixed, missing values must be filled in, and the data must be formatted so that it can be simply analyzed. Data fields that include NA, NaN, or Null or are empty are referred to as missing values. Missingness At Random (MAR), Missingness Not At Random (MNAR), and Missingness Completely At Random (MCAR) are all present in the aforementioned data set. If the above missing values do not affect the final results, they should be replaced with legitimate values, dummy values, or deleted.

Table: 2 Dataset with missing values

| **EPSON-PRINTER** | **ZEB-MONITOR** | **EPSON-INK** | **ZEB-MONITOR** | **NV-GRAPICS-CARD** | **349722.1** |
|---|---|---|---|---|---|
| **PENDRIVE** | DELL-LAPTOP-BAT | ANTI-VIRUS | | | 112729.2 |

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 11s***
***DOI: https://doi.org/10.17762/ijritcc.v11i11s.8175***
***Article Received:** 30 June 2023 **Revised:** 24 August 2023 **Accepted:** 06 September 2023*

| | | | | | |
|---|---|---|---|---|---|
| **SCREEN-GUARD** | EPSON-PRINTER | ANTI-VIRUS | MOUSE | KEYBOARD | 292318.1 |
| **EPSON-INK** | ZEB-MONITOR | EPSON-CATRIDG | ROUTER | EPSON-INK | 164656.1 |
| | ZEB-MONITOR | MOUSE | SCREEN-GUARD | EPSON-PRINTER | 58915.74 |
| **ZEB-MONITOR** | PENDRIVE | MOUSE | EPSON-CATRIDG | DELL-LAPTOP-BAT | 29217.45 |
| **EPSON-PRINTER** | | EPSON-CATRIDG | ZEB-MONITOR | | 129648.3 |
| **KEYBOARD** | UPS | SCREEN-GUARD | | UPS | 74696.09 |

The following missing values as shown in figure 2, from the above data set must be filled in with appropriate values. For improved outcomes, it is essential to create models that are trustworthy, valid, and high accuracy.
To get data into a format that can be analyzed, data transformation must be done. Normalisation, standardization, and discretization are common data transformation procedures. The data are scaled to a standard range via normalisation. For the CNAIS algorithm process, every item name that is a string must be encoded or turned into an integer.

## IV. RESULTS AND DISCSSIONS

**CNAIS Web Application** is web application for selecting dealer for special status with high sales of Non-Associated Items which is developed using HTML, JavaScript, CSS with Python at its back end for processing and the database Transaction-DB support is provided by MySQL Community server back end technology. Product business executives or marketing managers can use the CNAIS Web Application (CWA) to grant special status to dealers who have generated higher sales for their items. The CWA application is set up on a straightforward HTTP server, such as Apache Tomcat or XAMP. The user interface is straightforward, allowing users to enter individual sales data or upload sales data in an excel field or CSV file using the upload sales options. The selection option is utilised to view the outcomes of the chosen show rooms. CWA offers both tabular and clustered views.
The Figure 2 Shows cluster based on Associated Item and Non-Associated Items Purchase Amount.

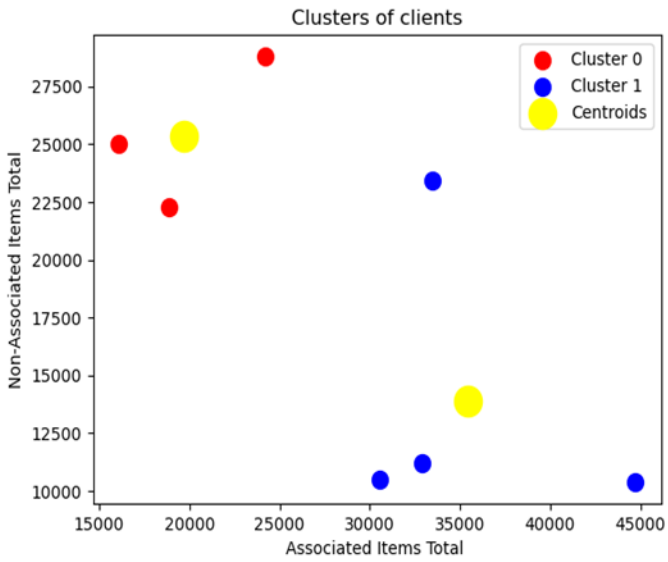


Figure 2. Cluster based on Associated Item and Non-Associated Items Purchase Amount.

The Figure 3 shows clusters based on Associated Item and Non-Associated Items Purchase Amount.

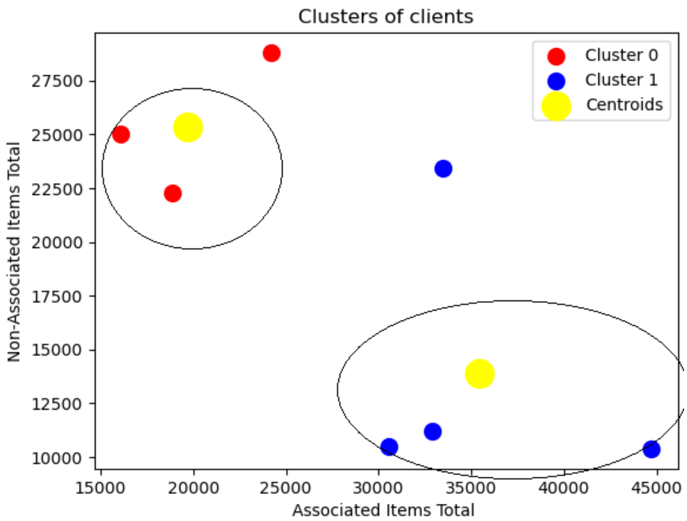


Figure 3. Clusters based on Associated Item and Non-Associated Items Purchase Amount.

The figure 4 below shows Threshold mark of considering Non-Associated Items Amount.

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 11s***
***DOI: https://doi.org/10.17762/ijritcc.v11i11s.8175***
***Article Received:** 30 June 2023 **Revised:** 24 August 2023 **Accepted:** 06 September 2023*

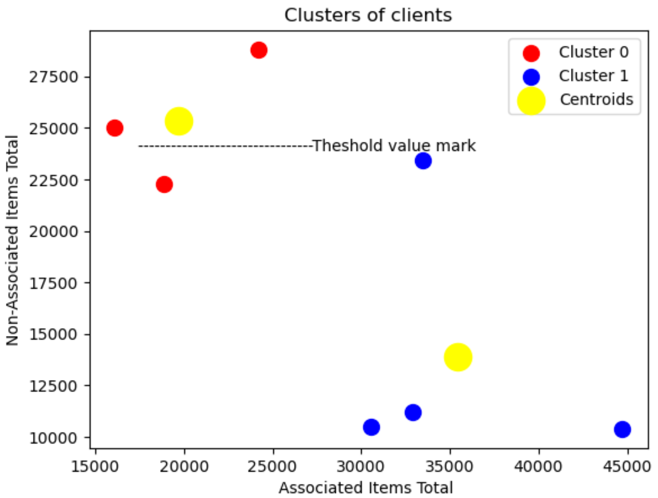


Figure 4. Threshold mark of considering Non-Associated Items Amount

The figure 5 below depicts the Threshold mark of considering Non-Associated Items Amount with Transactions Ids /Dealer Ids.

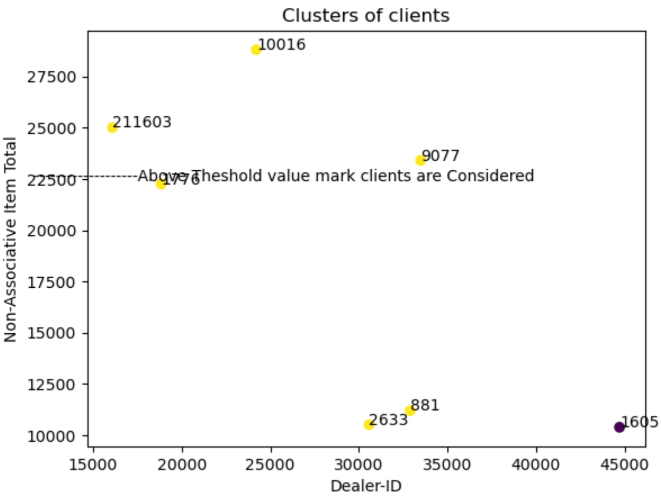


Figure 5. Threshold mark of considering Non-Associated Items Amount with Transactions Ids /Dealer Ids.

The list of dealers chosen for special status is given in the findings below (Fig 6), which were obtained by running an algorithm as a back-end service using data from a MySQL table or a CSV file. In the back servers-implemented and -run CNAIS Web application, the findings are shown graphically and in tabular format.

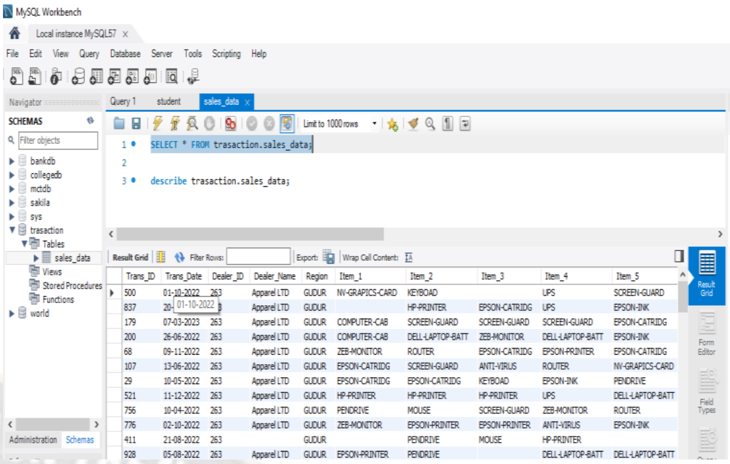


Figure : 6 Shows Table data uploaded into MYSQL database Table

## V. CONCLUSION

Applications connected to data science are presented in this paper before data mining. Association rule mining and clustering analysis's two major techniques are succinctly described, along with their applications. In order to implement a selection mechanism, the CNAIS Algorithm, which is a part of the research, is tested using real-time dealer sales data. Results are generated once the normalised data is tabulated in the data processing stage. Technical information is provided regarding the use of the CNAIS algorithm in web applications. Screens containing example data and results are displayed in the CNAIS Web Application (CWA), which has been built with easy interfaces.

**Data availability: The** data used to support the findings of this study will be shared by the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare that there is no conflict of interest regarding the publication of this paper.

## REFERENCES

[1] K.M.V. Madan Kumar,Dr. P.V.S. Srinivas, Algorithms for Mining Sequential Patterns International Journal of Information Sciences and Application. ISSN 0974-2255 Volume 3, Number 1 (2011), pp. 59-69

[2] Agrawal, R., Imielinski, T., and Swami, A. N. 1993. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, P. Buneman and S. Jajodia, Eds. Washington, D.C., 207–216.

[3] Peltier J W,Schibmwsky J A,Schuhz D E,et al. "Interactive Psychographics:Cross-Selling in the Banking Industry" .Journal of Advertising Research, 2002, 4 (2) ,pp.7-22.

[4] Sotiris Kotsiantis, Dimitris Kanellopoulos,Association Rules Mining: A Recent Overview , GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), 2006, pp. 71-82

[5] Bellini etal. “Multi Clustering Recommendation System for Fashion Retail” ,Multimedia Tools and Applications p1-28, 1573-7721 2022 Springer
[6] Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, Second Edition, Morgan Kaufmann Publishers is an imprint of Elsevier 2006.
[7] R. Agrawal and R. Srikant. Mining Sequential Patterns. In Proc. of the 11th Int’lConference on Data Engineering~ Taipei, Taiwan, March 1995.
[8] G.Ahalya and H. M. Pandey, “Data clustering approaches survey and analysis,” 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), 2015, pp. 532-537, doi: 10.1109/ABLAZE.2015.7154919.
[9] Rui Xu and D. Wunsch, “Survey of clustering algorithms,” in IEEE Transactions on Neural Networks, vol. 16, no. 3, pp. 645-678, May 2005, doi: 10.1109/TNN.2005.845141.
[10] Ramakrishna, M. T., Venkatesan, V. K., Bhardwaj, R., Bhatia, S., Rahmani, M. K. I., Lashari, S. A., & Alabdali, A. M. (2023). HCoF: Hybrid Collaborative Filtering Using Social and Semantic Suggestions for Friend Recommendation. Electronics, 12(6), 1365.
[11] Xu, D., Tian, Y. A Comprehensive Survey of Clustering Algorithms. Ann. Data. Sci. 2, 165–193 (2015). https://doi.org/10.1007/s40745-015-0040-1
[12] Prof. Amruta Bijwar, Prof. Madhuri Zambre. (2018). Voltage Protection and Harmonics Cancellation in Low Voltage Distribution Network. International Journal of New Practices in Management and Engineering, 7(04), 01 - 07. https://doi.org/10.17762/ijnpme.v7i04.68.
[13] Kanungo, Tapas, et al. “An efficient k-means clustering algorithm: Analysis and implementation.” Pattern Analysis and Machine Intelligence, IEEE Transactions on 24.7 (2002): 881-892.
[14] Shearer, C. (2000) The CRISP-DM Model: The New Blueprint for Data Mining. Journal of Data Warehousing, 5, 13-22.
[15] Raghunath, K. M. K. ., Kumar, V. V. ., Venkatesan, M. ., Singh, K. K. ., Mahesh, T. R. ., & Singh, A. . (2022). XGBoost Regression Classifier (XRC) Model for Cyber Attack Detection and Classification Using Inception V4. Journal of Web Engineering, 21(04), 1295–1322. https://doi.org/10.13052/jwe1540-9589.21413
[16] Mahesh, T. R., Ram, M. S., Ram, N. S. S., Gowtham, A., & Swamy, T. N. (2021). Real-Time Eye Blinking for Password Authentication. In Integrated Emerging Methods of Artificial Intelligence & Cloud Computing (pp. 428-434). Cham: Springer International Publishing.
[17] Christoph Schröer, Felix Kruse, Jorge Marx Gómez,A Systematic Literature Review on Applying CRISP-DM Process Model, Procedia Computer Science,Volume 181,2021,Pages 526-534,ISSN 1877-0509
[18] https://indatalabs.com/blog/why-python-is-used-in-data-science
[19] Mahesh, T.R., Vinoth Kumar, V., Shashikala, H.K., Roopashree, S. (2023). ML Algorithms for Providing Financial Security in Banking Sectors with the Prediction of Loan Risks. In: Sarveshwaran, V., Chen, J.IZ., Pelusi, D. (eds) Artificial Intelligence and Cyber Security in Industry 4.0. Advanced Technologies and Societal Change. Springer, Singapore. https://doi.org/10.1007/978-981-99-2115-7_14
[20] Devarajan, D., Alex, D. S., Mahesh, T. R., Kumar, V. V., Aluvalu, R., Maheswari, V. U., & Shitharth, S. (2022). Cervical cancer diagnosis using intelligent living behavior of artificial jellyfish optimized with artificial neural network. IEEE Access, 10, 126957-126968.
[21] B. N. Kumar, T. R. Mahesh, G. Geetha and S. Guluwadi, "Redefining Retinal Lesion Segmentation: A Quantum Leap With DL-UNet Enhanced Auto Encoder-Decoder for Fundus Image Analysis," in IEEE Access, vol. 11, pp. 70853-70864, 2023, doi: 10.1109/ACCESS.2023.3294443.
[22] Salim Mohammed Al-Waili, Zulkiflee Abd Latif, Siti Aekbal Salleh. (2023). GIS-Based Decision Support System and Analytical Hierrachical Process for Integrated Flood Management. International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 392–399. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2678.
[23] Mahesh, T.R., Vinoth Kumar, V., Shashikala, H.K., Roopashree, S. (2023). ML Algorithms for Providing Financial Security in Banking Sectors with the Prediction of Loan Risks. In: Sarveshwaran, V., Chen, J.IZ., Pelusi, D. (eds) Artificial Intelligence and Cyber Security in Industry 4.0. Advanced Technologies and Societal Change. Springer, Singapore. https://doi.org/10.1007/978-981-99-2115-7_14
[24] K. Gunasekaran, V. V. Kumar, A. C. Kaladevi, T. R. Mahesh, C. R. Bhat and K. Venkatesan, "Smart Decision-Making and Communication Strategy in Industrial Internet of Things," in IEEE Access, vol. 11, pp. 28222-28235, 2023, doi: 10.1109/ACCESS.2023.3258407.