International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 12 Issue 2, February-2023, Impact Factor: 7.957

Review Analysis on Tb Racks Dataset Using Data Mining Techniques

Mohini Dhummerkar[1], Dr. Neelesh Jain[2], Dr. Neeraj Gupta[3]

[1]M. Tech Research Scholar, Department Computer Science SAM College of Engineering and Technology
[2,3]Professor, Department Computer Science SAM College of Engineering and Technology

ABSTRACT

The objective of data mining is to extract appealing correlated information from bulky databases. This study proposal aims to comprehend the fundamental idea behind data mining technologies used in TB rack-based supermarket analysis. The Top to Bottom TB Racks Ratio (TBR)-based clustering technique is provided in a way that clarifies the use of data mining in supermarket analysis. The set of items that shoppers purchased at the grocery store were analyzed by the author using data mining software called Weka 3.8.6 for the TB racks application. The author attempted to relate the experiment and algorithm in this study. The author then presented the results by demonstrating a TB rack analysis application. The statistical outcome from Weka. The number of algorithms in the weka tool is used in this paper to analyze consumer data. We utilized the Random Tree algorithm and the BayesNet algorithm for classification in that algorithm. In the supermarket, consumer behavior analysis is important for making decisions because it can predict consumer behavior based on a variety of data. Consumer behavior analysis can also be used to find hidden relationships between data.

Keyword:-Supermarket, BayesNet, Randoom Tree, TB racks, Data mining, Weka.

INTRODUCTION

Companies are now able to collect massive amounts of data thanks to our highly technological age. Most organizations of businesses have amassed tens of thousands to hundreds of millions of pieces of data that, if not converted into useful information, have no value[1]. The technology known as data mining is a tool that can be used by businesses to extract data from large databases. Knowledge discovery is a broader process that includes data mining.

A lot of people think of data mining as just one part of a larger process called Knowledge Discovery in Databases (KDD).As per Fayyad et.al, 'KDD is the nontrivial cycle of recognizing substantial, novel, possibly helpful and at last justifiable examples in information[6].'

RELATED WORK

Trnka A. et al 2010[1] presented how to apply Six Sigma methodology to Market Basket Analysis. Information Mining techniques give a ton of chances on the lookout area. One of them is a basket market analysis. Numerous statistical techniques are utilized in the Six Sigma methodology. We can alter the process's Sigma performance level and improve outcomes by incorporating Market Basket Analysis into one of Six Sigma's phases as part of Data Mining.

Chenyang M. et al 2019[2] Describe Nowadays, many people use data mining to find connections between items in huge datasets. Frequent itemset mining is an essential component of association rule mining, one of the most popular data mining techniques. The trustworthy but curious cloud service provider (CSP) receives large amounts of data. As a result, the CSP and third parties must be prevented from obtaining the raw data. In addition, supermarket transactions are too few to be mined using the same techniques as the majority of the other data. If these methods are applied to this particular dataset, they will require more computational power than they would for standard data. Under the encrypted mining query on supermarket transactions, we present an effective protocol to determine whether an item set is frequent. We develop a blocking algorithm to enhance mining efficiency. This algorithm reduces the mining process's computation cost by separating encrypted transactions into blocks and only calculating bilinear pairings on ciphertexts of part blocks rather than all ciphertexts. Finally, we conduct theoretical

analyses and simulator experiments to assess the efficiency of our protocol in terms of correctness, running time, cost of computation, and security. Our protocol clearly outperforms the previous solution in terms of efficiency while maintaining the same level of security, as demonstrated by the results.

Riccardo G. et al 2019 [3] Explains nowadays, offering personalized services to customers is a major challenge for supermarket chains. One of these services is market basket prediction, which provides customers with a shopping list for their next purchase based on their current requirements. The various factors that influence a customer's decision-making process cannot be simultaneously captured by current methods: co-occurrence, regularity, and recurrence of the purchased goods. We define a Temporal Annotated Recurring Sequence (TARS) pattern that can simultaneously and adaptively capture all of these factors to achieve this goal. TARS Based Predictor (TBP) is a predictor for the next basket that, in addition to TARS, is able to comprehend the level of the customer's stocks and recommend the set of the most essential items. We also define the method for extracting TARS. Supermarket chains could effectively expedite their customers' shopping sessions by implementing the TBP, which would allow them to create individualized recommendations for each customer. Extensive testing demonstrates that TBP outperforms the most recent competitors and that TARS is capable of explaining customer purchase behavior.

RESEARCH PROBLEM STATEMENT

The majority of businesses today do not recognize the significance of data mining techniques for the organizations' benefit. The study of shopping baskets has grown in popularity among retailers in recent years. They were able to collect data on their clients and their purchases thanks to cutting-edge technology. The use and application of transactional data in supermarket analysis increased with the introduction of electronic point-in sales[4]. Analyzing this kind of data is extremely helpful in retail businesses for comprehending buyer behavior. Mining buying designs permits retailers to change advancements, and store settings and serve clients better.

This is probably because hundreds of organizational-related software and tools have appeared in supermarkets, causing many corporate employees to become confused. As a result, the purpose of this study would be to investigate the significance of data mining to the organization, both directly and indirectly.

METHODOLOGY

Quantitative, qualitative, and demand research are the three main types of research strategies. Both experimental and non-experimental types of research are possible. The purpose of this subchapter is to investigate the TBR-based method of clustering. A set of data items will be divided into appropriate groups using the clustering algorithm. A collection of transactions serves as the data representation in the TB racks-based supermarket analysis. Product codes are represented by rows and columns in this dataset [5]. A “Yes/No” value indicates whether that product was purchased during that transaction in each cell.

Purpose of Research

The purpose of the proposed study is to examine how shopping patterns are related to rack layout and promotion in a sample supermarket store by analyzing customer purchases. The purpose of this study is to determine the outcomes of implementing TB racks analysis in a supermarket. The project's goals are as follows:

- To investigate the concept of the clustering algorithm in order to investigate the fundamental idea of datamining technology.
- To use a data mining tool known as weka 3.8.6 to carry out the application of TB racks–based supermarketanalysis to discover hidden patterns among various supermarket products.
- The TB racks experiment aims to determine which of these products sells well together so that when a customergoes grocery shopping, the related products can be arranged together to increase the likelihood of a sale.

RESEARCH METHODOLOGY

There are a variety of relevant and utilized research methods in information systems research. Problems that haven't been studied before can benefit from exploratory research. It will aid in problem definition and comprehension. There will not be conclusive outcomes using this method. In this instance, the research will begin with a broad idea, and the findings can be applied to subsequent studies

Purpose of Research

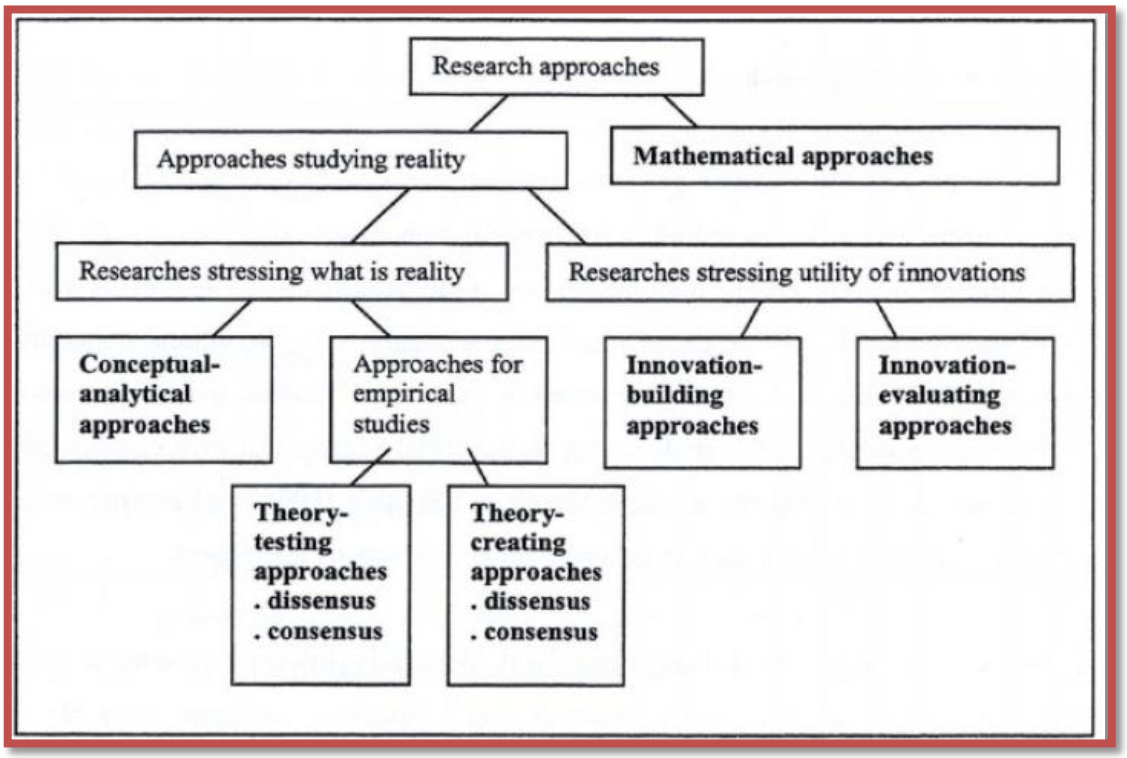


Figure 1:- Taxonomy of research methods.

Techniques used in the Experiment;
First, a definition of the work's problem, current issues, and current art was used to conduct the research. The next step is to collect the data that will be used in the research[22].

- A number of data sets have been utilized. These data sets were chosen with the diversity of data and data type in mind, which will be taken into account when determining the effects of the test's methodologies. A large part of the various informational collections are utilized top incorporates a large part of the variety with the information types; including nominal, integer, and categorical data, as well as a combination of these types in some cases.
- The distribution of the data within a set, also known as the normal distribution or even distribution (the distribution describes how the data instances are distributed among the classes in the entire dataset: The rationed distribution of instances across all classes is referred to as the normal distribution, whereas skew data refers to data that is split between a small number of classes and not across all classes. The distribution graph and the skew data can be used to measure the skew.
- In order to alter the various outcomes of the tools applied to various data sets and data types, classification algorithms have been utilized. First, these algorithms are explained in great detail. Each tool's outcome is checked and analyzed using the algos' output. Taking into account all of the parameters used to verify the resulting output, this result is used to verify the tool's capability and accuracy[12].
- The tools used are among the most widely used and well-liked by users: KNIME, WEKA, SQL, and Hadoop Tanagra They are free internet-accessible open-source tools. The various capabilities of each tool for data manipulation, representation, and other features are taken into consideration when selecting these tools. In the proposed research, the Weka tool was favored.

Tools for Data Mining: a description of the data mining techniques that were utilized in this comparative study.

Weka:
Weka offers a data mining system package with all of the features needed to carry out the data mining process. This includes a very effective use of every classifier algorithm that has been determined thus far. This is very useful and can be used in a lot of different ways [13]. It is a Java-based tool that lets the user work on the service using a command prompt or a graphical user interface shown in figure 2.

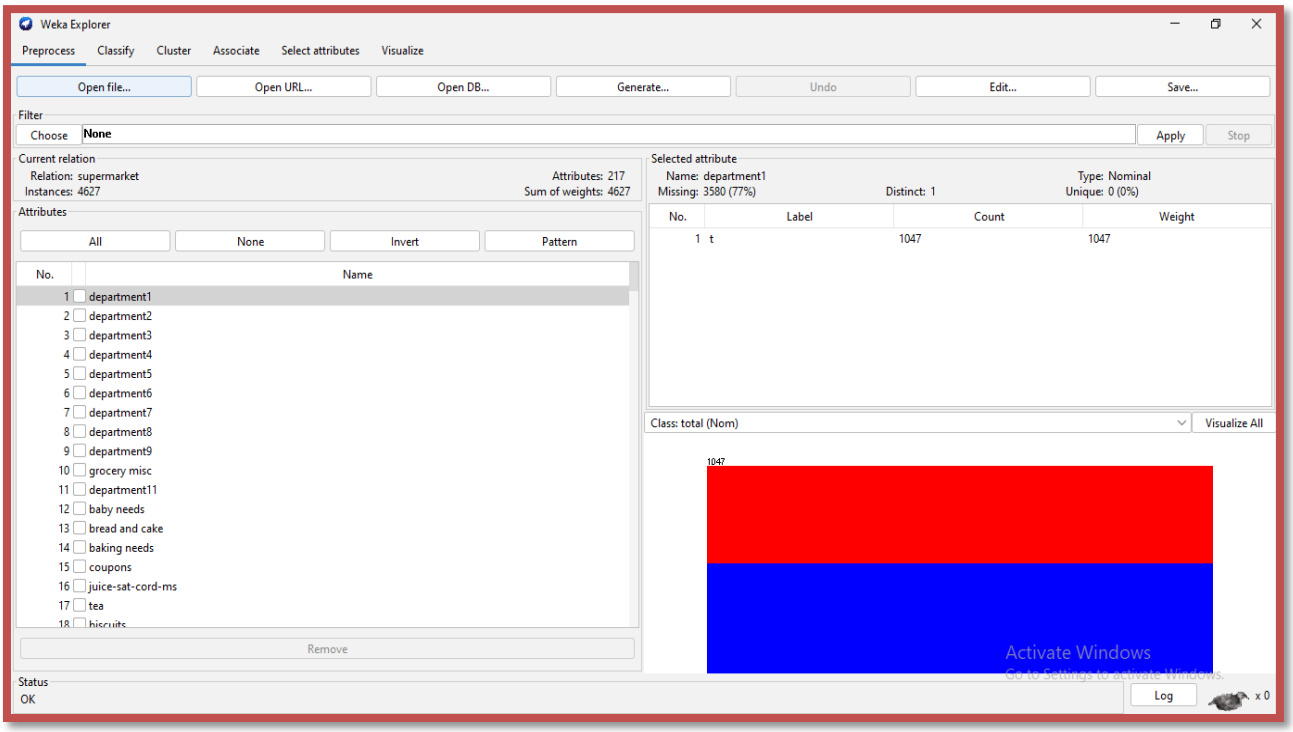


Figure 2:- Weka Tool explorer Supermarket Interface

Evaluation parameters for the tool's performance include:
It is not correct to evaluate the performance of the classifying algorithm solely by looking at the value of accuracy achieved by any classifier. The value of the instances classified as belonging to their actual class is all that matters to the classifier's accuracy[16]. It does not introduce the classifier's additional features, such as; the relationship between the attributes of the data, the measure of the correct distribution of data instances to each and every class that is conceivable, the number of positive outcomes from all of the positive outcomes that are received, and a number of other things[23].

Association rules:
Supermarket analysis frequently makes use of association analysis, which is also known as affinity analysis or association rule mining. ARM is currently the best method for analyzing big supermarket data. However, when there are a lot of products in a lot of sales, the data matrix for association rule mining typically becomes large and sparse, which takes longer to process. This kind of information is provided by association rules in the form of "IF-THEN" statements. To comprehend an association rule's presence, nature, and strength, three common indexes are used. 2004, Berry and Linoff; 2005, Larose; Zhang and Zhang, 2002)

Lift is acquired first since it gives data on regardless of whether an affiliation exists or on the other hand assuming the affiliation is positive or negative. We get the value for support if the value for lift indicates that there is an association rule[17].

$$Left = \frac{S(X \cap Y)}{S(X) \times S(Y)} \tag{3.1}$$

Support of an item or item_set is the fraction of transactions in our dataset that contain that item or item_set. Because it's possible for a rule with low support to simply happen by chance, this is a crucial measure. From a business perspective, a low support rule may also be boring because it may not be profitable to promote items that are rarely purchased together. As a result, uninteresting rules are frequently eliminated using support.

$$support = \frac{S(X \cap Y)}{N} \tag{3.2}$$

Confidence is defined as the conditional probability that shows that the transaction containing the LHS will also contain RHS. Interpretation of association analysis results should be cautious. The association rules' 16 inference does not necessarily imply causality. Instead, it suggests that the items in the rule's antecedent and consequence are strongly related to each other.

$$Confidence = \frac{S(X \cap Y)}{S(X)} \quad (3.3)$$

Confidence and support measure the strength of an association rule.. Due to the large size of the transactional database, there is a greater chance that we will end up with too many irrelevant rules. Prior to the analysis, we typically establish a threshold of support and confidence to ensure that our result contains only interesting and useful rules.

If lift is greater than 1, it indicates that the likelihood that the items on the RHS will occur in this transaction has increased as a result of their presence on the LHS. If the lift is less than 1, it means that the likelihood that the items on the RHS will be included in the transaction is lower because they are on the LHS. If the lift is 1, it indicates that items on the LHS and RHS are independent of one another: The likelihood that items will occur on the RHS is unaffected by the fact that it is known that the items on the LHS are present[18].

When we conduct supermarket analysis, we search for rules with multiple lifts. Rules with high confidence are those where the probability of an item appearing on the RHS is high, given the presence of items on the LHS. Rules with high support are also preferable because they will be applicable to a large number of transactions.

Confusion matrix, Recall, Accuracy:
A confusion matrix is a kind of table that shows how many instances of data are actually classified and how many are incorrectly classified. The number of classes defined for the data set is denoted by the number "n" in this n×n matrix.

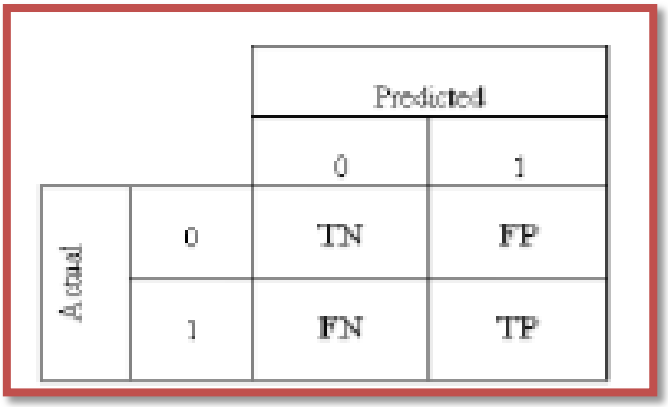


Figure 3:- confusion matrix

This confusion matrix fills in as the reason for determining values for pretty much every other boundary utilized. The classifier's predicted values are shown in the columns, while the data object's actual values or class labels are shown in the rows[20].

The cell values are as follows:
True negative: Which proportion of the negative cases were correctly classified?
Calculated as: ***TN/(TN+FP)***

False Positive: Which proportion of negative cases were incorrectly classified as positive?
Calculated as: ***FP/(TN+FP)***

False Negative: Which proportion of positive cases were incorrectly classified as negative?
Calculated as: ***FN/(FN+TP)***

True Positive or Recall: proportion of correctly classified positive cases.
Calculated as: ***TP/(FN+TP)***

Accuracy : it signifies the quantity of right expectations made by the classifier. And the figure 3.3 is as follows:
(TN+TP)/(TN+FP+FN+TP

Precision (Confidence): *Precision is the percentage of positives that were predicted to occur but actually occurred. The following is an indication of this value:* TP/(FP+TP)

Transaction:
An agreement, contract, understanding, or transfer of cash or property that imposes a legal obligation on two parties is known as a transaction. Events that start a change in the asset, liability, or net worth account in accounting terms.

First, transactions are recorded in the journal and then added to the ledger. The profit and loss account, balance sheet, and other accounting books follow next.

A transaction in banking is an action taken by an account holder on their own initiative that has an effect on a bank account[8].

A transaction is an exchange of goods or services between a buyer and a seller in the language of commerce. It has three parts:
(1). Money and goods are transferred,
(2). Possession may or may not be transferred in conjunction with title transfer,
(3). Right transfer in exchange.

A typical transaction in marketing involves a customer purchasing a set of products from a retail store or online. All of the data entered into the database about each individual transaction can be found in these transactions. These can include information about the customer, the products purchased, the time of purchase, and whether or not the company's marketing strategies are attracting customers, among other things.

Additionally, a transaction can occur simultaneously or over a longer period of time, such as a day, quarter, or fiscal year. because they can be used for any occasion.

Rule generation in Apriori algorithm
This section will describe the Supermarket Analysis algorithm that will be running behind the Python libraries. Companies will be able to better understand their customers and analyze their data with greater precision as a result of this. The first associative algorithm, the Apriori algorithm, was used in subsequent advancements of association, classification, and associative classification algorithms[19].

The Apriori algorithm employs a level-wise approach to generate association rules, with each level representing the number of items in the rule's consequent. In the beginning, all high-confidence rules with only one item are extracted. From these existing rules, new ones are then created.
For example, if:
{a, c, d}→ {b}
{a, b, d} →{c}

Are high-confidence rules, then the candidate rule {a, d} → {b, c} is created by combining the results of the two rules. Analyses of associations: Principles and guideline**s**).

To put it another way, the generation of a candidate rule involves combining two rules that have the same prefix in the rule consequent [7].
Association rule mining is thought to involve two steps:

Frequently Creating Itemsets:
Find all incessant thing sets with help not entirely settled least help count. In most cases, interesting associations and correlations between item sets in relational and transactional databases are discovered through frequent mining. In a nutshell, Frequent Mining reveals the items that are associated with a transaction or relationship. Multiple iterations are required to find frequent item sets. Scan the entire training data to count new candidate item sets from existing item sets. In a nutshell, it only requires two significant steps:
1. Pruning
2. Joining

Making regulations:
Prepare a list of all frequent item-set association rules. Determine the support and confidence for each rule. Rules that fall short of the minimum support and confidence thresholds should be pruned.

Frequent Itemset Generation
It is searches the entire database for the frequent itemset that meets a support threshold. It is the step with the highest computational cost because it scans the entire database. In the real world, retail transaction data can exceed gigabytes and terabytes, necessitating the use of an optimized algorithm to eliminate item sets that will not aid in subsequent steps. The Apriori algorithm is used for this.

Apriori calculation satisfies "Any subset of an incessant itemset should likewise be regular. In other words, there is no requirement to generate or test a superset of an uncommon itemset.

The Apriori algorithm principle is depicted graphically in the image below. It consists of a relation between subsets

of the k-item-set and the k-item-set node. Figure 4 shows that you start with the null set and work your way up, creating subsets, until you reach the entire transaction data.

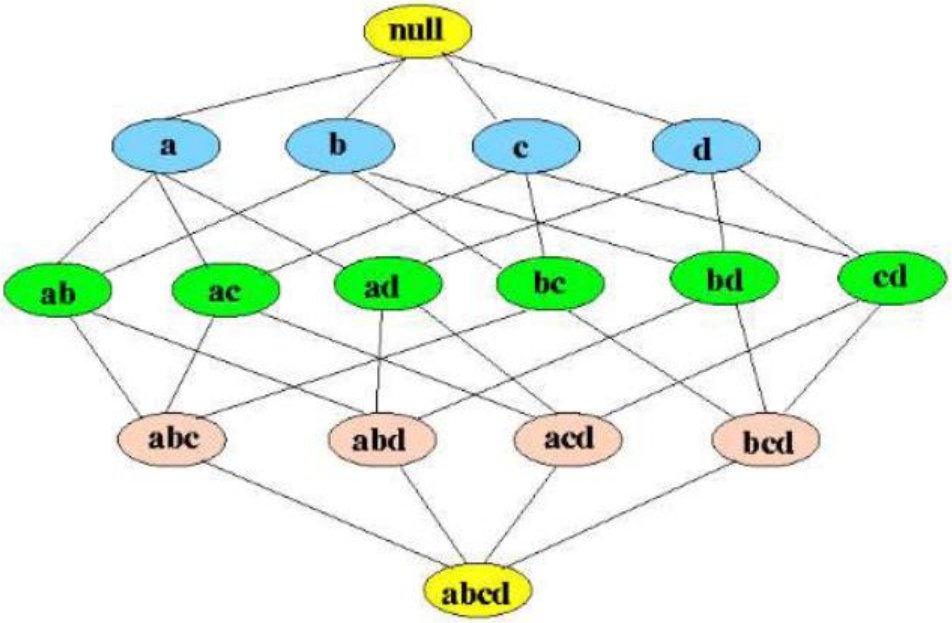


Figure 4: All Possible Subsets

This demonstrates shown in figure 3. that finding support for each combination will be difficult to generate frequent item-sets. As a result, the Apriori algorithm contributes to reducing the number of sets that must be generated, as shown in the figure below

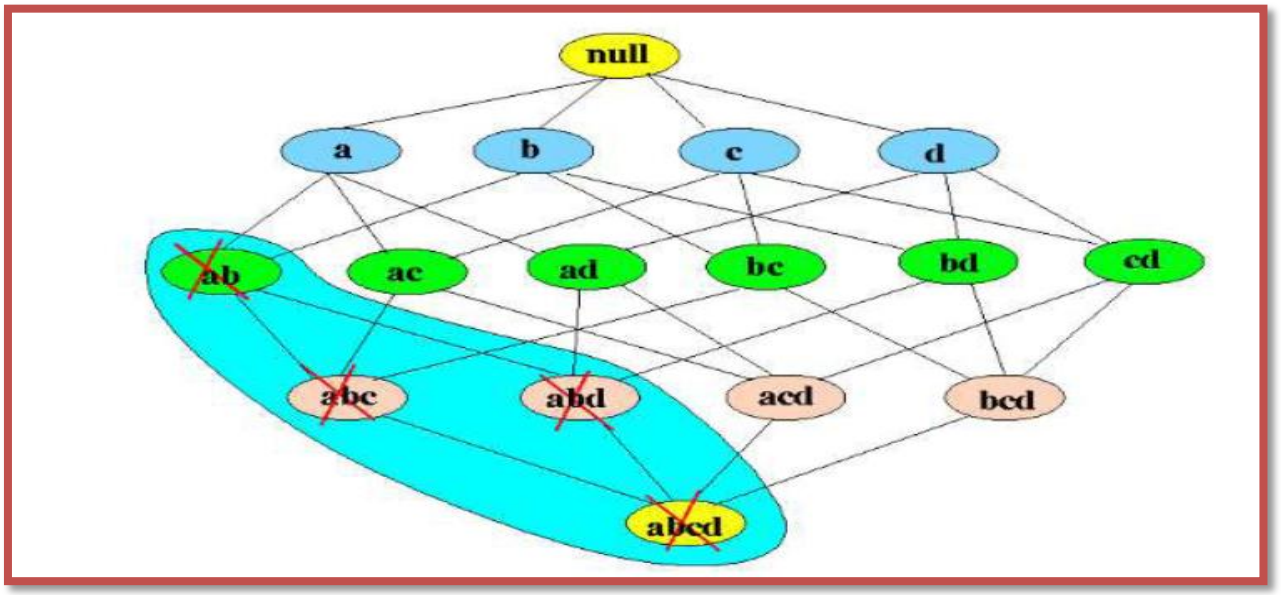
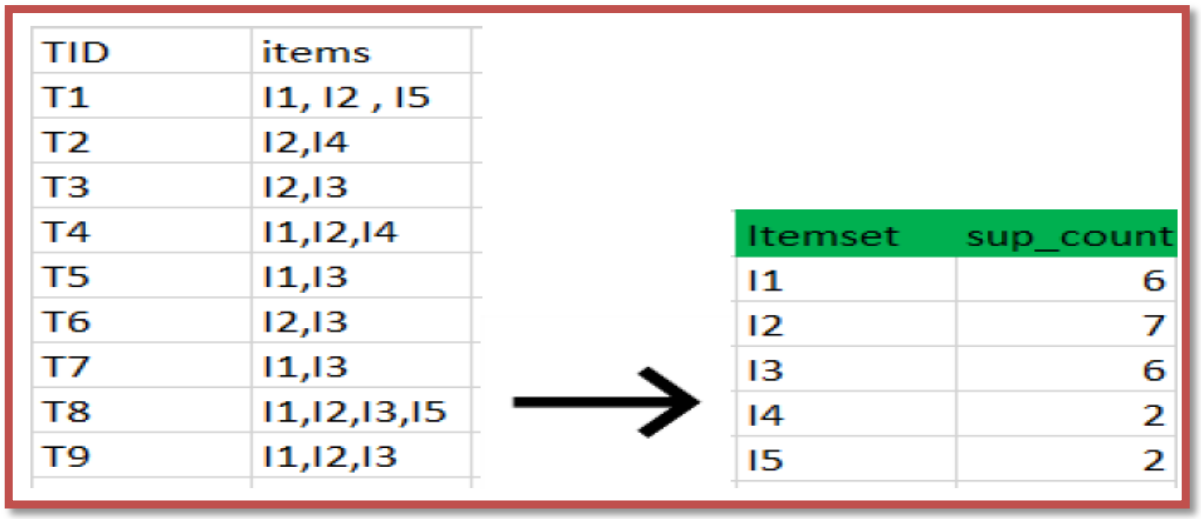


Figure 5: If an Item Set Is Infrequent, We Do Not Consider Its Super Sets

We need not take into account all of an item set's super sets if it is uncommon. It can also be examined as a transactional data set. You can see why the Apriori algorithm is so much more effective and produces stronger association rules step by step in the following example.

Step: 1.

1. Prepare a table with the support count for each dataset item.
2. Compare the support count to the minimum support count (in this instance, the minimum support count is 2); if the support count is lower than the minimum support count, those items should be taken out, and a new set of items will be produced[24].



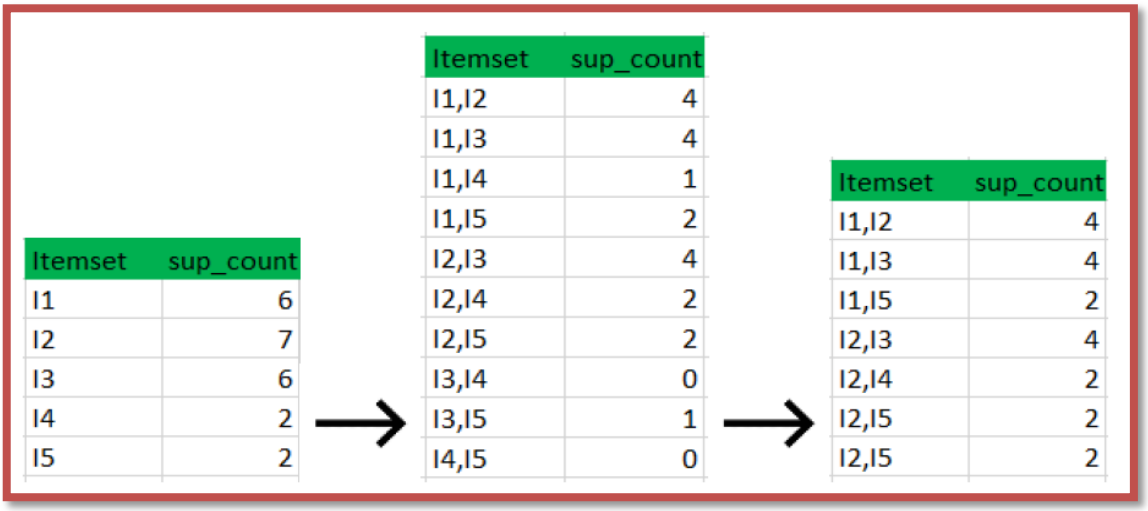
TID	items
T1	I1, I2 , I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

Itemset	sup_count
I1	6
I2	7
I3	6
I4	2
I5	2

Figure 6: Transactional Data To Frequent Items

Step 2:

1. The join step is the name of this step. By joining all of the items together in a cross-joining, we create a new set.
2. Determine whether an itemset's subsets are common or not, and if not, remove that itemset. For instance, we can see that the subset of "I1, I2" is frequent and is called "I1" and "I2." We must perform identical checks on each itemset.
3. Now, search the dataset to determine the item-set support count.
4. Since we have already established a minimum support count threshold of 2, We compare the minimum support count and remove those items if the support count is lower than the minimum support count. provides us with an additional itemset, as shown below[25].



Itemset	sup_count
I1	6
I2	7
I3	6
I4	2
I5	2

Itemset	sup_count
I1,I2	4
I1,I3	4
I1,I4	1
I1,I5	2
I2,I3	4
I2,I4	2
I2,I5	2
I3,I4	0
I3,I5	1
I4,I5	0

Itemset	sup_count
I1,I2	4
I1,I3	4
I1,I5	2
I2,I3	4
I2,I4	2
I2,I5	2
I2,I5	2

Figure 7: Pruning And Joining

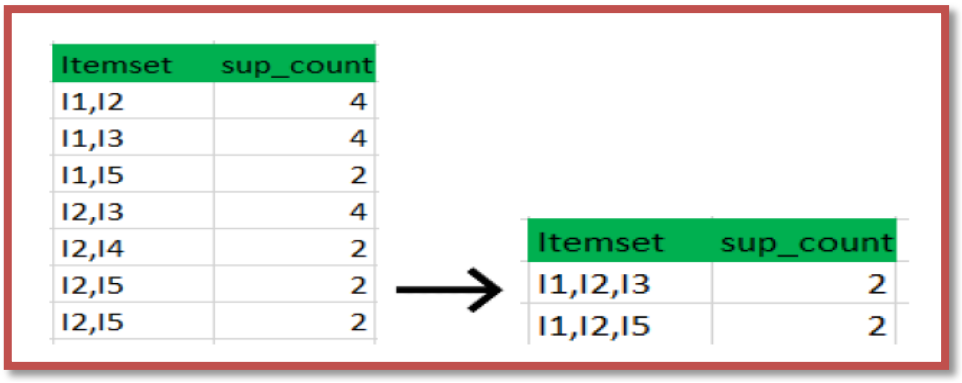
Step 3:

1. Following the acquisition of a second dataset, the join step remains unchanged. We link each itemset together crosswise. As a result, the generated itemset will be:

{I1, I2, I3}
{I1, I2, I4}
{I1, I2, I5}
{I1, I3, I5}
{I2, I3, I4}
{I2, I4, I5}
{I2, I3, I5}

2. Verify whether each subset of these item sets is frequent; if not, remove that subset.
3. For instance, the frequent subset of {I1, I2, I3} in this instance are {I1, I2}, {I1, I3},and {I2, I3}. However, one of the subsets for {I2, I3, I4} is {I3, I4},, which is uncommon. So, we get rid of this. Every itemset is treated in the same way[26].
4. After removing all non-frequent item sets, search the dataset to determine the support count for the remaining item set.

5. If the support count is lower than the minimum support count, then those items should be taken out. As can be seen below, it provides us with another itemset.



Itemset	sup_count
I1,I2	4
I1,I3	4
I1,I5	2
I2,I3	4
I2,I4	2
I2,I5	2
I2,I5	2

Itemset	sup_count
I1,I2,I3	2
I1,I2,I5	2

Figure 8: Pruning And Joining again until there are no more frequent items left

Step 4:

1. Once more, we follow the same procedure. In the first place, we do the join step and we cross join each itemset with each other. In our example, the item set's first two components ought to match.
2. After that, determine whether or not each subset of these item sets is frequent. The itemset formed following the join step in our example is called {I1, I2, I3, I5}. Therefore, the uncommon {I1, I3, I5} subset of this itemset is one of its subsets. As a result, there is no longer an itemset.
3. We come to a stop here because no more itemsets are found frequently.

The first step in association rule mining was this.

The next thing to do is make a list of all the frequently used item sets and figure out how strong the association rules are. We do this by determining the confidence of each rule. The following formula is used to determine our level of confidence[27]:

$$Rule: X \Rightarrow Y$$

$$Support = \frac{frq(X,Y)}{N}$$

$$Confidence = \frac{frq(X,Y)}{frq(X)}$$

$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

We will demonstrate the process of rule generation by using the example of any frequent item (I1, I2, I3):

Rules	Formula	Confidence
{I1, I2} => {I3}	2 / 4 * 100	50,00%
{I1, I3} => {I2}	2 / 4 * 100	50,00%
{I2, I3} => {I1}	2 / 4 * 100	50,00%
{I1} => {I2, I3}	2 / 6 * 100	33,33%
{I2} => {I1, I3}	2 / 7 * 100	28,57%
{I3} => {I1, I2}	2 / 6 * 100	33,33%

Figure 9: Calculation of confidence

Therefore, the first three rules can be considered strong association rules in this instance if the minimum confidence is fifty percent. Take, for instance, **{I1, I2} => {I3}** where a confidence level of 50% indicates that 50 percent of people who bought **"I1" and "I2"** also bought **"I3."**

Cluster Analysis

Data mining and analysis applications rely heavily on clustering. In data mining, cluster analysis is a fundamental method for identifying similar objects in a large database. The term "cluster" refers to a collection of data objects that are similar to one another within the same cluster but distinct from those in other clusters. The basic structure of a cluster is an ordered list of data with similar characteristics. No matter how they are shaped, cluster analysis can identify them. It is a method of unsupervised learning. Scalability, the capacity to deal with noisy data, and imperceptibility to the order of the input records are the most important requirements for clustering algorithms[14].

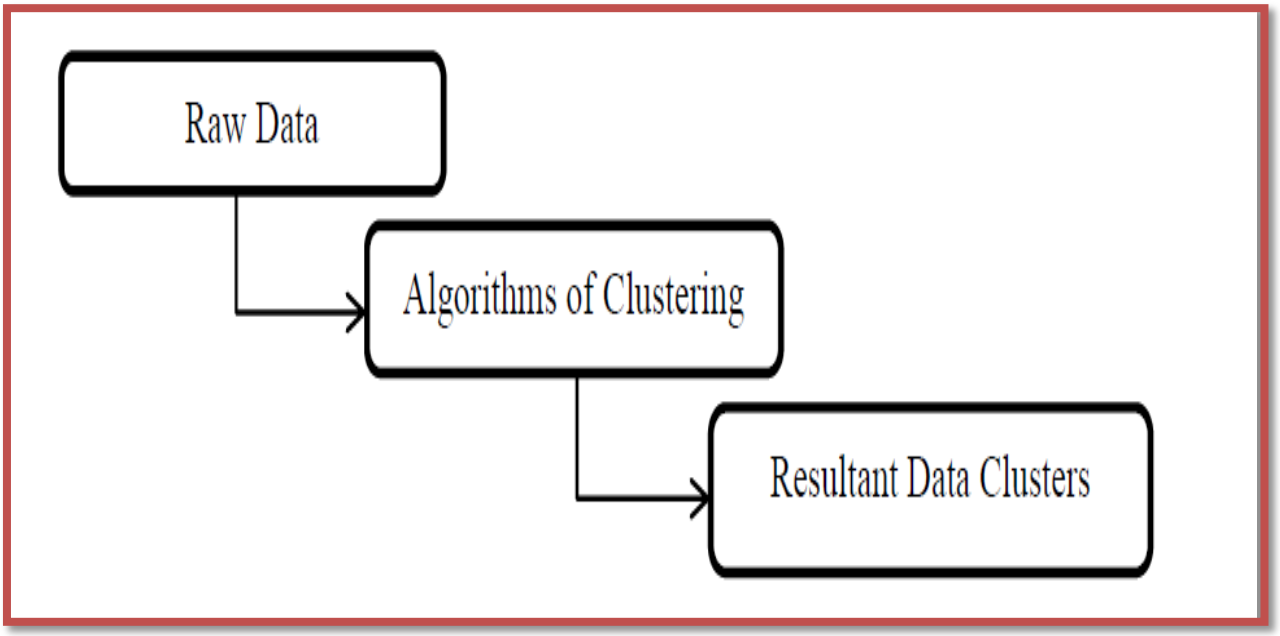


Figure 10:- Stages of Cluster Analysis

Clustering plays an important role, as from a practical point of view it is used in various data mining applications such as marketing, CRM, medical diagnostic, information retrieval and text mining, web analysis and many others. It is a machine learning technique used to put similar data elements into related groups without having any prior knowledge of group definitions.

Proposed algorithms:

For frequent item set mining, the most popular algorithm is the proposed Apriori. It begins by locating the frequently occurring individual items in the transactional database before expanding them to ever-larger itemsets until they become sufficiently prevalent in the database[15].

When there are no more extensions that meet the minimum support condition, the algorithm ends.

The algorithm's main idea is to search the database for itemsets that are frequently used, then prune items that are found to be less frequently used in subsequent steps. The join and prune steps in candidate generation are extremely significant. C_{k+1} is produced in the first step when R_k is joined to itself. Since it cannot be a subset of the frequent (k+1) itemset, any infrequent k-itemsets are pruned during the prune step.

The Proposed Apriori algorithm can be represented in the following steps:

C_k – candidate itemsets with size k;

R_k – frequent itemsets with size k.

1. Add the most frequent items to Lk (k=1)
2. A collection of candidate itemsets Ck+1 of size (k+1) can be created using Lk.
3. Perform a database search to identify the frequently occurring items in Ck+1 and move them into R_k+1.
4. If R_k+1 is full:

K:=k+1; proceed to step No.

In a few easy steps, the Apriori works as shown in the example below. Let's say that the following sets make up a sample transaction database: a, c, d, b, c, e, a, b, c, e, b, e, and so on. There is a product in the assortment for each letter. For instance {a} is cleanser, {b} is hair conditioner.

The first step involves the algorithm adding up the frequencies of each item's individual supports. The minimum support level can be predefined if we want to guarantee that an item is frequent. The minimum support required here is 2. As a result, four of the items are discovered to be common.

All of the 2-pairs of frequently occurring items are compiled in the subsequent step. For the purpose of further investigation, the already discovered uncommon items are excluded[16]. The Apriori algorithm reduces the number of possible combinations to find all possible two-item pairs.

A list of all three-triplets of frequent items is produced in the final step by connecting a frequent pair to a frequent single item. The algorithm comes to an end at this point because the pair of four items generated at the following step does not have enough support.

Integration With Supermarket Analysis

Based on the literature review, it seems most logical to start with the Apriori algorithm because it is the most widely used and easiest to use in R. Once the general architecture is built, it is fairly simple to run various analyses and algorithms on the data. Accessing the necessary data to run the Apriori algorithm in TB rack ratio is simple when using the dimensional model as a source of analysis[17]. Since R scripts can be used within stored procedures, the proposed architecture is based on Weka 3.8.6 tool and later. Figure 3.11 provides an illustration of the proposed architecture for the TB rack ratio integration.

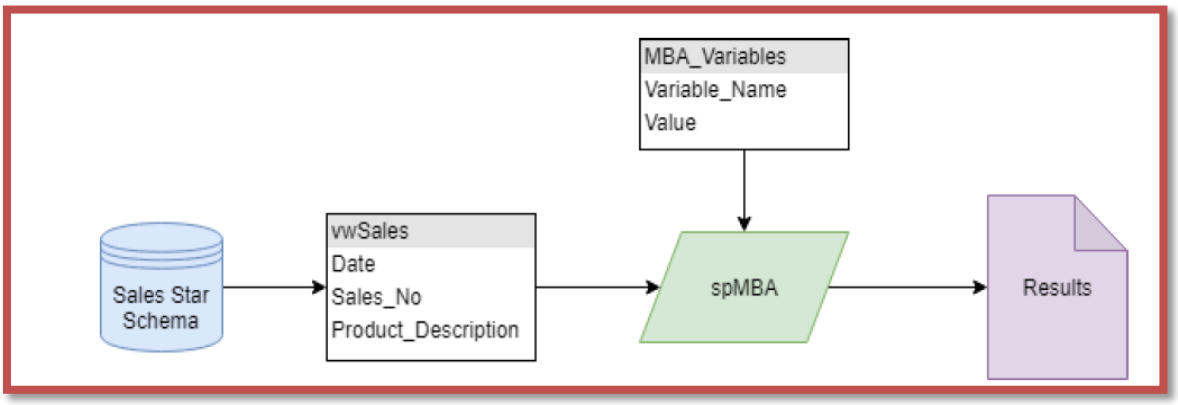


Figure 11: Proposed TB rack integration.

Data Arrangement and analysis Supermarket

The data must be fed to the Weka software in order to find product affinities. Since the available data is in CSV format, the first step is to drag and drop a CSV node from the source tab. The Excel dialog box that allows you to browse, open, and preview the source data can be accessed by double-clicking on the node.

Applying the Apriori algorithm to the data is the next step. The Apriori node is selected from the modeling tab for this purpose. Support is set to 1 by default, and confidence is set to 40 by default; these numbers must be tested later. In the majority of association rule mining programs, the maximum number of antecedents is set to six by default.

There would only be two or three items associated logically. As a result, the default setting will be used. Accept this support and confidence for the time being, and then click Apply. The connection between the first and second nodes follows. The run button can finally be pressed. A new node with the shape of a gem is added and connected to the other nodes automatically when the process is run.At this point, the affinity analysis's findings can be seen by double-clicking the gem node. However, keep in mind that because the program had difficulty processing Farsi product names, the data that was sent to the software only contained the product code. As a result, the results will contain the code; however, you will need to manually check the code in the product code table.

CONCLUSION

There are a number of aspects of retailing that would be difficult without supermarket analysis. Product tracking not only aids in the management and processing of inventory, but it can also be used in cross-sale campaigns and promotional strategies because it provides an overview of co-occurrence products. Marketers and retailers are informed of a possible influential strategy that can affect the sale of one or both of the products when they know which ones are more likely to be chosen together.

The classification approach is a supervised learning algorithm that is used in data mining. It identified the categorized data, allowing for the predetermined classification. One of the most important studies in data mining is

the data classification problem, in which the minimum classification of the data of interest is used and very little rack bottom sample data is used in comparison to rack top classes. Because this causes classifier prediction to be based on the majority class, solutions to this issue must be found. On the consumer behavior dataset, we used WEKA to evaluate solutions to class imbalance issues. In this paper, we contrast the two classification algorithms. This analysis aims to determine that the Random Tree algorithm produces fewer accurate classified data than the Bayes Net algorithm.

REFERENCES

[1]. Trnka Andrej, "Market Basket Analysis with Data Mining Methods Six Sigma methodology improvement", International Conference on Networking and Information Technology, pp 446-449, IEEE 2010.
[2]. Chenyang Ma, Baocang Wang , Kyle Jooste, Zhili Zhang , and Yuan Ping, "Practical Privacy-Preserving Frequent Itemset Mining on Supermarket Transactions" IEEE SYSTEMS JOURNAL Personal use is permitted, but republication/redistribution requires IEEE permission, pp-1-11 , IEEE 2019.
[3]. Riccardo Guidotti, Giulio Rossetti , Luca Pappalardo , Fosca Giannotti, and Dino Pedreschi, "Personalized Market Basket Prediction with Temporal Annotated Recurring Sequences", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 31, NO. 11, pp- 2151-2163 November 2019.
[4]. A. M. Khattak, A. M. Khan, Sungyoung Lee and Young-Koo Lee. "Analyzing Association Rule Mining and Clustering on Sales Day Data with XLMiner and Weka", International Journal of Database Theory and Application Vol. 3, No. 1.
[5]. Andreas Mild, Thomas Reutterer, "An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data", Journal of Retailing and Consumer Services, Volume 10, 123-133, 2003.
[6]. David R. Bell and James M. Lattin, "Shopping Behavior and Consumer Preference for Store Price Format: Why "Large Basket", Marketing Science, Vol. 17, No. 1, 66-88, 2008.
[7]. Kumar, N., & Rao, R., "Using Basket Composition Data for Intelligent Supermarket Pricing" . Marketing Science, 25(2), 188-199,2006.
[8]. Ayinde A. Adetunji A., Bello M., and Odeniyi O., "Presentation evaluation of naive bayes and decision stump algorithm s in mining students educational data.," Interna-tional Journal of Computer Science Issues (IJCSI), vol. 10, no. 4, 2013.
[9]. Joachims T., Freitag D., and Mitchell T., "Web-watcher: A tour guide for the world wide web," in IJCAI (1), Citeseer, 1997, pp. 770–777,.
[10]. Romero C. and Ventura S., "Educational data mining: a review of the state of the art, Systems, Man, and Cybernetics, Part C: Applications and Reviews," IEEE 2010 Transactions on, vol. 40, no. 6, pp. 601–618.
[11]. Mendes R. R., Voznika F. B. de, Freitas , and Nievola J. C., "Discovering fuzzy classification rules with genetic programming and co-evolution," Principles of Data mining and Knowledge Discovery , Springer, 2001, pp. 314–325,.
[12]. Zhang B., Legible Discovering And Readable Chinese Typefaces For Reading Digital Documents. PhD thesis, Concordia University, 2011.