**Project Report**

1. **Introduction**

Market Basket Analysis most popularly known as Association Rules. Most business organizations produce huge data from their daily transactions. Similarly, customer purchase data is collected in grocery stores. But not all the information is useful. It is important to extract useful information from the given data. The process of extracting useful information is called data mining. This process involves steps like selection, preprocessing, transformation, data mining, and interpretation. The retailers are interested in finding the purchase pattern of the customers. This will help the retailers to analyze business-related data and make decisions in the organization.

The Association analysis is very much helpful in finding the hidden patterns in the large datasets. These patterns or relations are represented in the form of association rules or frequent item sets. The association analysis can also be applied in bioinformatics, medical diagnosis, web mining, and scientific data analysis.

The market basket data can be represented in binary format. Considering the transactional data, each row is corresponding to the transaction and the column refers to the items.

1. **Literature Survey**

Data mining has taken an important part in marketing literature in the last few decades. Market basket analysis is also one of the oldest areas of data mining and is most commonly used for mining association rules. Various algorithms are developed by researchers in the field of Association Rule Mining (ARM) and Clustering to help users achieve their objectives.

The market Basket Analysis technique is used to find the items that are bought together and how buying one product influence to purchase another product. MBA is a tool, which can keep track of buying pattern of customers.

Two main researchers named Ramakrishnan Srikant and Rakesh Agrawal developed the apriori algorithm for finding frequent patterns in large datasets and analyzes the customers buying the products and the percentage of total sales of the product. Such association rules are helpful to find the leading products.

* 1. **Existing System**

The existing algorithms work on static data and they do not capture changes in data with time. But proposed algorithm not only mines static data but also provides a new way to take into account changes happening in data. This will be helpful to examine the customer behavior and assist in increasing the sales.

**Disadvantages of existing system**

The existing system records the data in ledgers or Excel sheets.

Customer data is not maintained properly.

Does not simulate any report.

Does not identify frequent item set.

* 1. **Proposed System**

The proposed system works on Association Rules and Apriori algorithm to mine the data.

Catalog design

Cross marketing

Association rule mining

Association Rule Mining is used to identify the association between items from the dataset. Association analysis is helpful for discovering relations present in large data set. The rule says that a strong relationship should exist between the items. There are two issues to be discussed are discovering the patterns in the large data set and some of the patterns may be spurious because they may occur by chance.

The association rule is expressed as X → Y, where X and Y are taken as disjoint item sets

(i.e. X ∩ Y = ∅). The strength of association rule can be expressed using Support and Confidence.

Support: -

The number of transactions that contain all the items over the total number transactions.

If support value is high, it means that items are more frequently to occur.

The support gives an idea that how many times an item set has occurred in the overall transactions.

Support ⇔ 𝑆𝑢𝑝𝑝𝑜𝑟𝑡 𝐴 =

Confidence:-

Confidence is a measure of the likelihood that customer buy product A will buy product B as well. A rule of association is therefore a remark of the form (item set A) ⇒ (item set B) where

A is the precedent and B is the consequence. Confidence gives the probability of Consequence

Occurring on the cart provided with pre-existing antecedents. For frequently appearing Consequent, it doesn’t matter what the customer have it in the Antecedent. The confidence of an Association rule, which results very often, will always be of greater value.

⇔

Confidence 𝐴 =

Lift

Given that different items are bought at different frequencies, how do we know that 2-items really do have a strong association.

There are multiple ways to express the formula to calculate lift. Let me first show what the formulas look like, and then I will describe an intuitive way for you to think about it.

Lift(A→B) = Probability(A & B) / (Support(A) \* Support(B)

Lift(A→B) = Confidence(A & B) / Support(B)

lift>1 means that the two items are more likely to be bought together, while lift<1 means that the two items are more likely to be bought separately. Finally, lift=1 means that there is no association between the two items.

Conviction

Conviction is another way of measuring association, although it is a bit harder to get your head around. It compares the probability that A appears without B if they were independent with the actual frequency of the appearance of A without B. Let’s take a look at the general formula first:

**Conviction (A→B)**= (1 - Support (B)) / (1 - Confidence (A→B))

**2.3 Feasibility Study**

**2.4 Tools and technologies used**

**3. System Requirements**

**4. Software requirements specifications**

**5. System Design**

**5.1 Data Flow Diagram**

**5.2 Use case Diagram**

**5.3 Activity Diagram**

**6. Implementation**

**6.1 Module1**

**6.2 Module2**

**7. Testing**

**7.1 Unit Testing**

**7.2 Functionality Testing**

**7.3 Integration Testing**

**7.4 Verification and Validation Testing**

**8. Future Enhancements**

**9. Conclusion**

At present many data mining algorithms have been developed and applied on variety of practical problems. However periodic mining is a new approach in data mining which has gained its significance these days. This field is evolving due to needs in different applications and limitations of data mining. This would enhance the power of existing data mining techniques. Finding out the patterns due to changes in data is in itself an interesting area to be explored. It may helpful in x Find out interesting patterns from large amount of data. x Automatically track the changes in facts from previous data; due to this feature it may be helpful in fraud detection. x Predicting future association rules as well as gives us right methodology to find out outliers. Authors suggested that, some areas are still there which need to be focused on. Firstly, results have influenced greatly by the manual threshold values for score, so it is needed to automate the threshold values for better recognition of outliers. Secondly, this approach is specifically targeted at Market Basket Data, it may perhaps be extended to other areas.

**10. Bibliography**

<https://www.sciencedirect.com/science/article/pii/S1877050916305208>

<https://www.diva-portal.org/smash/get/diva2:1031325/FULLTEXT02>

<https://towardsdatascience.com/association-rules-2-aa9a77241654>

<http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/apriori/>

<https://towardsdatascience.com/apriori-algorithm-for-association-rule-learning-how-to-find-clear-links-between-transactions-bf7ebc22cf0a>

<https://www.researchgate.net/publication/343484851_Market_Basket_Analysis_Recommendation_System_Using_Association_Rules>