Market Basket Analysis

Kartik Saini, Khushi Sharma, Akshaj Agarwal

Mentor: Ms. Anjali Sardana

*Computer Science Engineering-Artificial Intelligence Machine learning Department ABES Engineering College*

[Khushi.20B1531039@abes.ac.in](mailto:Khushi.20B1531039@abes.ac.in)

[Akshaj.20b1531044@abes.ac.in](mailto:Akshaj.20b1531044@abes.ac.in)

[Kartik.20B1531055@abes.ac.in](mailto:Kartik.20B1531055@abes.ac.in)

Abstract**— Data analysis plays a vital role in the present era as it helps us to understand patterns by exploring them in meaningful ways. Market basket is one of the main methods used to find frequently occurring items in a transactional database. As these databases are large, it is impossible to read the database manually and extract the pattern. Market Basket Analysis is one of the main techniques of data mining that focuses on finding the purchase patterns of the customers.**

**Market Basket Analysis will be implemented by using the Apriori algorithm. This algorithm determines the frequent data or item set from the transaction database.**

**The problem many retailers face is the placement of the items. They are unaware of the purchasing habits of the customer.**

**Market basket analysis can help the shop managers determine the strong relationships between the items which ultimately helps them to put products that co-occur together close to one another.**

Keywords— Transactions, Database, Apriori, Market Basket, Data Analysis

1. Introduction

Market basket analysis is that technology that is used to analyze the items/goods in one or more shopping baskets that a customer has in one particular moment.

Market basket analysis application ought to be designed and implemented at a supermarket not only owing to being able to help the sales promotion design but also able to be made as a reference to re-manage item stock’ incoming and outcoming in the warehouse.

Market basket analysis is a [data mining](https://searchsqlserver.techtarget.com/definition/data-mining) technique used by retailers to increase sales by better understanding customer purchasing patterns. It involves analyzing large data sets, such as purchase history, to reveal product groupings, as well as products that are likely to be purchased together.

Implementation of market basket analysis requires a background in statistics and [data science](https://www.techtarget.com/searchenterpriseai/definition/data-science), as well as some algorithmic computer programming skills.

There are two types of market basket analysis:

1. Predictive market basket analysis: This type considers items purchased in sequence to determine cross-sell.
2. Differential market basket analysis: This type considers data across different stores, as well as purchases from different customer groups during different times of the day, month or year. If a rule holds in one dimension (like store, time period or customer group), but does not hold in the others, analysts can determine the factors responsible for the exception. These insights can lead to new product offers that drive higher sales.

In market basket analysis, [association rules](https://www.techtarget.com/searchbusinessanalytics/definition/association-rules-in-data-mining) are used to predict the likelihood of products being purchased together. Association rules count the frequency of items that occur together, seeking to find associations that occur far more often than expected.

Algorithms that use association rules include AIS, SETM and Apriori. The Apriori algorithm is commonly cited by data scientists in research articles about market basket analysis and is used to identify frequent items in the database, then evaluate their frequency as the datasets are expanded to larger sizes.

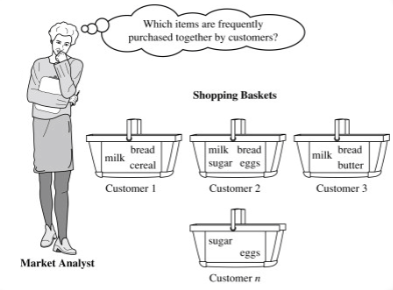


Fig i: Customer purchase pattern

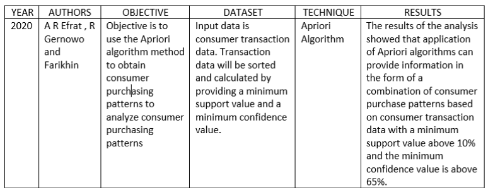
1. PROBLEM STATEMENT

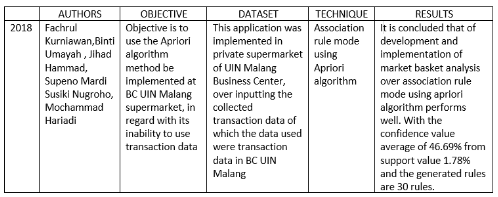
# The problem many retailers face is the placement of the items. They are unaware of the purchasing habits of the customer.

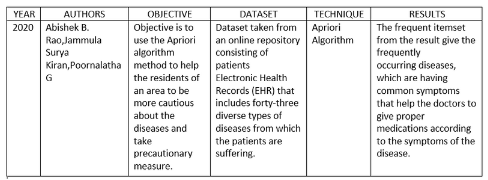
# Market basket analysis can help the shop managers determine the strong relationships between the items which ultimately helps them to put products that co-occur together close to one another.

1. LITERATURE REVIEW

The related work associated with our project is given below by the different research papers read whose summary is as follows:

Research Paper 1:

Research Paper 2:

Research Paper 3:

1. PROJECT OBJECTIVE

To analyze the Brazilian E-Commerce Public Dataset by Olist, taken from Kaggle. The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil.

Its features allows viewing an order from multiple dimensions: from order status, price, payment, product attributes.

To implement an algorithm best suited for the market basket analysis which would be apriori algorithm. Market Basket Analysis will be implemented by using the Apriori algorithm. This algorithm determines the frequent data or item set from the transaction database.

To perform market basket analysis using apriori algorithm and find out the strong relationships between the items which ultimately helps shopkeepers to put products that co-occur together close to one another.

1. PROPOSED METHODOLOGY

The proposed methodology related to our project is given below:

Step 1: Choosing the appropriate dataset. Here we will use Brazilian E-Commerce Public Dataset by Olist

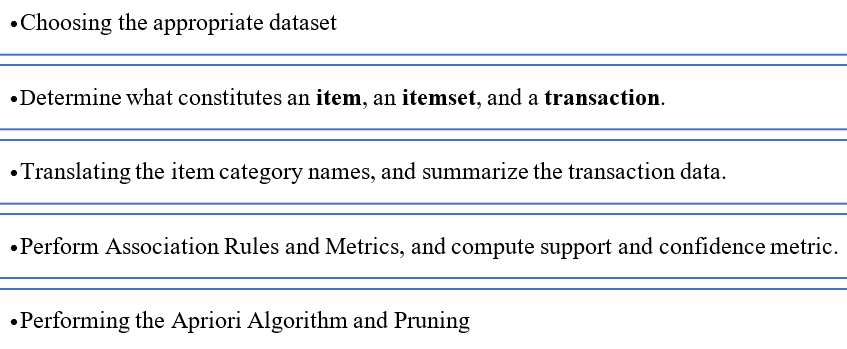
Step 2: Determine what constitutes an item, an itemset, and a transaction. This will depend on the dataset we're using and the question we're attempting to answer.

Step 3: We will translate the item category names, Convert product IDs to product category names,   
Construct transactions from order and product data and Summarize final transaction data.

Step 4:We now perform Association Rules and Metrics, One-hot encode the transaction data, Compute the support metric, Compute the item count distribution over transactions, Create a column for an itemset with multiple items, Aggregate the dataset further by combining product sub-categories

Compute the confidence metric.

Step 5:Performing the Apriori Algorithm and Pruning, Computing association rules from Apriori output in the form of table given below, The leverage metric, Visualizing patterns in metrics.



1. DESIGN AND IMPLEMENTATION

The design of our project would be:

**Fig.3.** Work Flow Diagram

We will first choose the appropriate dataset. Here we will use Brazilian E-Commerce Public Dataset by Olist.

We then determine what constitutes an item, an itemset, and a transaction. This will depend on the dataset we're using and the question we're attempting to answer.

We will translate the item category names, Convert product IDs to product category names,   
Construct transactions from order and product data and Summarize final transaction data.

We now perform Association Rules and Metrics, One-hot encode the transaction data, Compute the support metric, Compute the item count distribution over transactions, Create a column for an itemset with multiple items, Aggregate the dataset further by combining product sub-categories.

Performing the Apriori Algorithm and Pruning, Computing association rules from Apriori output in the form of table given below, The leverage metric, Visualizing patterns in metrics.

1. RESULTS AND DISCUSSION

The result we got from analyzing the transaction is given below in Fig.3.

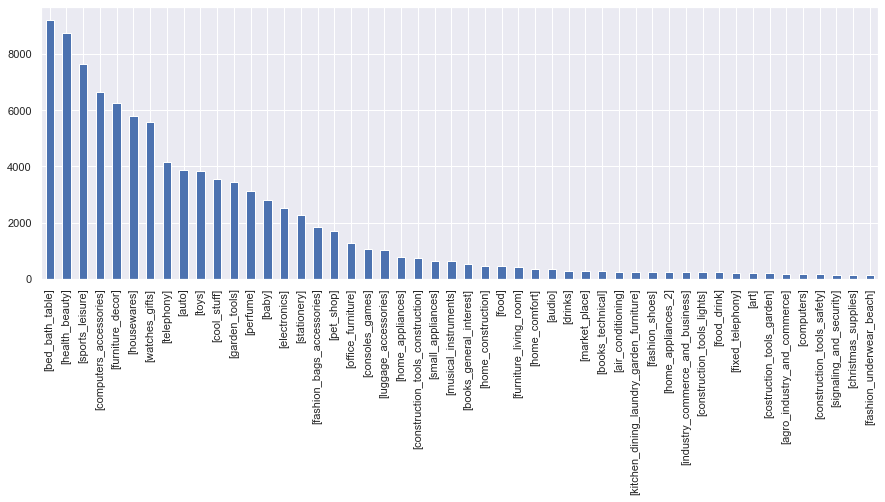


Fig.4. Plot of 50 largest categories of transactions

Fig.3. shows that the maximum category used in the transaction is [bed\_bath\_table} with more than 8000+ transactions followed by [health\_beauty] , [sports\_leisure] , etc.

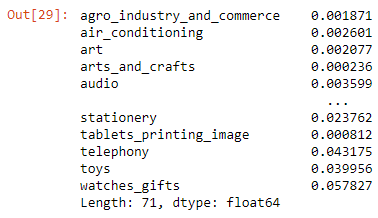


Fig.5. Computed the support metric

1. CONCLUSION AND FUTURE SCOPE

The project will give the overall market basket analysis using apriori algorithm and find out the strong relationships between the items which ultimately helps shopkeepers to put products that co-occur together close to one another.

We implemented the algorithm best suited for the market basket analysis which would was apriori algorithm. This algorithm determined the frequent data or item set from the transaction database.

The main disadvantage of the apriori algorithm is that it is slow due to:

1.A large number of itemsets in the Apriori algorithm dataset.

2.It scans the database many times.

That’s why the algorithm becomes inefficient and slow with large databases.

To solve this problem we would be performing

1.Transaction Reduction/Pruning - This method reduces the number of transactions scanned in iterations. The transactions which do not contain frequent items are marked or removed.

2.Using new database mapping way to avoid scanning the database repeatedly

3.Using overlap strategy to count support to achieve high efficiency.

Acknowledgment

The heading of the Acknowledgment section and the References section must not be numbered.

1. REFERENCES

[1] A R Efrat1, R Gernowo and Farikhin. Information System Department, Faculty of Postgraduate School, Diponegoro University, Semarang, Indonesia Physics Department, Faculty of Mathematics and Natural Sciences, Diponegoro University, Semarang, Indonesia , Mathematics Department, Faculty of Mathematics and Natural Sciences, Diponegoro University, Semarang, Indonesia: Consumer purchase patterns based on market basket analysis using apriori algorithms.

[2] Fachrul Kurniawan , Binti Umayah, Jihad Hammad, Supeno Mardi Susiki Nugroho, Mochammad Hariadi: Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data

[3] Abishek B. Rao, Jammula Surya Kiran, Poornalatha G: Application of market–basket analysis on healthcare.

[4] Kaggle: https://www.kaggle.com/olistbr/brazilian-ecommerce

ACKNOWLEDGMENT

We would like to convey our sincere thanks to **Ms. Anjali Sardana** for giving the motivation, knowledge and support throughout the course of the project. The continuous support helps in a successful completion of project. The knowledge provided is very useful for us.

We also like to give a special thanks to the department of CSE-AIML for giving us the continuous support and opportunities for fulfilling our mini project.

.