صورة تحتوي على نص, شعار, لقطة شاشة, التصميم

تم إنشاء الوصف تلقائياً



**Task 1: Market Basket Analysis**

**"Instacart Dataset"**

**DS3114))**

**Dr. Omaima Fallath))**

|  |  |
| --- | --- |
| **ID:** | **Name:** |
| **444005403** | **Sara Ayed Alsehli** |
| **443009106** | **Lina Ibrahim Alghamdi** |

**Table of Contents:**

[**Introduction 3**](#_Toc180843207)

[Overview Dataset: 3](#_Toc180843208)

[What it Represents: 4](#_Toc180843209)

[**Data Processing 5**](#_Toc180843210)

[Data Import: 5](#_Toc180843211)

[Data Merging: 5](#_Toc180843212)

[Data Filtering: 5](#_Toc180843213)

[Basket Representation: 5](#_Toc180843214)

[Transformation for Apriori: 5](#_Toc180843215)

[**Rules Discovered and Their Potential Applications 6**](#_Toc180843216)

[Rule Example: 6](#_Toc180843217)

[Real-World Application: 6](#_Toc180843218)

[**Conclusion 7**](#_Toc180843219)

[**References 8**](#_Toc180843220)

Introduction

The Instacart Market Basket Dataset contains information on customer orders from Instacart. It includes details about the products, aisles, departments, and customer order histories. The objective of this analysis is to discover patterns in customer purchasing behavior using association rule mining. By understanding these patterns, we can derive valuable insights for product recommendations and marketing strategies**.**

## **Overview Dataset:**

The Instacart Market Basket Analysis dataset consists of several files that capture various aspects of customer orders. Here's a breakdown:

**Number of Rows:** The main dataset contains millions of rows, each representing a product purchased in an order. For example, one file includes order\_products\_\_train.csv with over **3 million rows.**

**Number of Columns:** The dataset typically has **4 to 6 columns**, with important ones such as: Dataset Description:

* order\_id: Identifies each unique order.
* product\_id: Represents the ID of each product in the order.
* user\_id: Identifies the customer who placed the order.
* add\_to\_cart\_order: Shows the sequence in which items were added to the cart.
* reordered: Indicates if the item was previously ordered by the customer (1 if yes, 0 if no).
* product\_name, aisle\_id, department\_id: Provide product details, such as the product name and the department or aisle where the product belongs.

## **What it Represents:**

**Orders:** Information about the order history of customers, including what products were bought, the sequence of items, and whether they were reordered.

**Products:** Details about the products purchased, categorized by aisle and department.

**Customer Behavior:** By analyzing the orders, it is possible to detect purchasing patterns, frequent item combinations, and customer preferences.

# 

# **Data Processing**

## **Data Import:**

Multiple CSV files containing products, orders, aisles, and departments were loaded into Pandas (DataFrames). This allowed us to access all necessary information, such as which products were bought in each order.

## **Data Merging:**

We merged the relevant datasets to obtain a comprehensive view of orders, products, and customer behavior. This was done to enable effective pattern mining across multiple features.

## **Data Filtering:**

The dataset was filtered to focus on the top **15,000** products and orders to make the dataset more manageable and to focus on the most frequent items.

## **Basket Representation:**

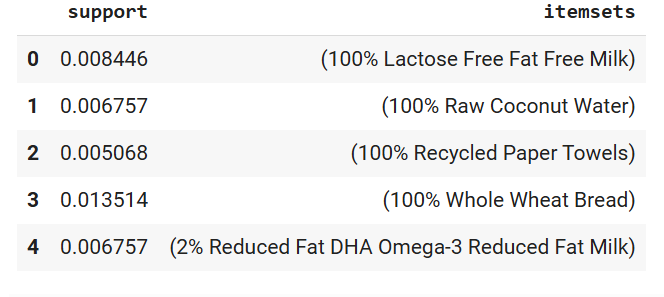
The data was transformed into a basket format and removed null value where each row represented a unique order, and the products purchased in that order. This format is required for association rule mining.

## **Transformation for Apriori:**

Using the (TransactionEncoder), the basket data was transformed into the format required for running the Apriori algorithm. This step converts each order into a binary matrix where each product is represented as a separate column.

# 

# **Rules Discovered and Their Potential Applications**

Using the **(Apriori algorithm**), frequent itemsets were generated, followed by the extraction of association rules based on metrics such as support, confidence, and lift. The most notable rules discovered include:

## **Rule Example:**

Customers who buy All Purpose Flour are also likely to buy Bananas. This rule had a confidence of 75% and a high lift value, indicating a strong association between these products.

## **Real-World Application:**

This information can be used to improve product recommendations. For example, when a customer adds All Purpose Flour to their cart, they can be recommended Bananas as a frequently bought together item.

# **Conclusion**

This analysis uncovered several useful purchasing patterns within the Instacart dataset. By applying these rules, Instacart or similar retailers could enhance their product recommendation systems and improve customer experience through personalized marketing campaigns. These patterns offer actionable insights into customer behavior and can guide inventory management and cross-selling strategies.

# **References**

(<https://www.kaggle.com/c/instacart-market-basket-analysis/data>)