MARKET BASKET INSIGHTS

Phase4

Project Introduction :

* This project aims to analyze market basket data. In this project, we used association rules and apriori algorithm to train our model.

Association Rules:

Association rule learning is a type of unsupervised learning technique that checks for the dependency of one data item on another data item and maps accordingly so that it can be more profitable. It tries to find some interesting relations or associations among the variables of dataset. It is based on different rules to discover the interesting relations between variables in the database.

The association rule learning is one of the very important concepts of machine learning, and we employed this rule in Market Basket analysis.

Association rule learning can be divided into three types of algorithms:

1. **Apriori**
2. **Eclat**
3. **F-P Growth Algorithm**

Among these algorithm we mainly used apriori algorithm for market basket insights

Apriori Algorithm:

* This algorithm uses frequent datasets to generate association rules. It is designed to work on the databases that contain transactions.
* This algorithm uses a breadth-first search and Hash Tree to calculate the item set efficiently.
* It is mainly used for market basket analysis and helps to understand the products that can be bought together

The steps for the apriori algorithm:

Step-1: Determine the support of itemsets in the transactional database, and select the minimum support and confidence.

Step-2: Take all supports in the transaction with higher support value than the minimum or selected support value.

Step-3: Find all the rules of these subsets that have higher confidence value than the threshold or minimum confidence.

Step-4: Sort the rules as the decreasing order of lift.

Dataset used: <https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis>

Code:

 import pandas as pd

import numpy as np

from mlxtend.frequent\_patterns import apriori, association\_rules

import matplotlib.pyplot as plt

import seaborn as sns

import re

import networkx as nx

import pandas as pd

dataset\_path = '/kaggle/input/market-basket-analysis/Assignment-1\_Data.xlsx'

df = pd.read\_excel(dataset\_path)

print("Number of rows and columns:", df.shape)

print("\nData Types and Missing Values:")

print([df.info](http://df.info/" \t "_blank)())

print("\nFirst few rows of the dataset:")

print(df.head())

basket\_plus = (df.groupby(['Transaction ID', 'Product'])['Unique Purchases']

               .sum().unstack().reset\_index().fillna(0).set\_index('Transaction ID'))

basket\_plus.drop('(not set)', axis=1, inplace=True)

basket\_plus

def encode\_units(x):

    if x <= 0:

        return 0

    if x >= 1:

        return 1

    basket\_encode\_plus = basket\_plus.applymap(encode\_units)Filtering The Transaction(Bought More Than 1 Items Only)

basket\_filter\_plus = basket\_encode\_plus[(basket\_encode\_plus > 0).sum(axis=1) >= 2]

basket\_filter\_plus

frequent\_itemsets\_plus = apriori(basket\_filter\_plus, min\_support=0.03, use\_colnames=True).sort\_values('support',ascending=False)

reset\_index(drop=True)

frequent\_itemsets\_plus['length'] = frequent\_itemsets\_plus['itemsets'].apply(lambda x: len(x))

df\_encoded = pd.read\_csv('transaction\_data\_encoded.csv')

from mlxtend.frequent\_patterns import apriori, association\_rules

frequent\_itemsets = apriori(df\_encoded, min\_support=0.007, use\_colnames=True)

rules = association\_rules(frequent\_itemsets, metric="confidence", min\_threshold=0.5)

print("Association Rules:")

print(rules.head())