IBM PHASE 2- PROJECT SUBMISSION

NAME : SAMEERA BANU M

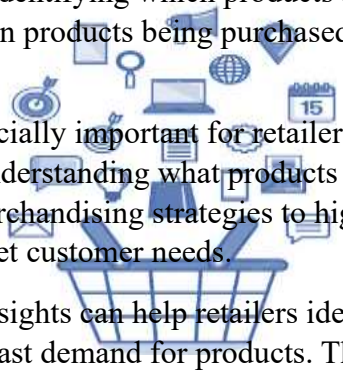
TITLE : MARKET BASKET INSIGHTS

DOMAIN : ARTIFICIAL INTELLIGENCE (AI)

PHASE 2 – INNOVATION

INTRODUCTION:

Market basket insights refer to the analysis of purchase patterns and behaviour of customers in a retail setting. This analysis involves identifying which products are commonly purchased together, understanding the frequency of certain products being purchased, and identifying any trends or patterns in customer behaviour.



Market basket insights are especially important for retailers who want to increase sales and improve customer satisfaction. By understanding what products are most commonly purchased together, retailers can adjust their merchandising strategies to highlight complimentary products or create product bundles that better meet customer needs.

Furthermore, market basket insights can help retailers identify the most profitable products, optimize pricing strategies, and forecast demand for products. This data-driven approach to retail management can help retailers optimize their inventory levels, reduce waste, and improve their bottom line.

Overall, market basket insights provide a powerful tool for retailers to better understand their customer and improve the efficiency, profit, sales which is applicable both in online and offline stores.

DATA SOURCE:

DATASET LINK: ***https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis***

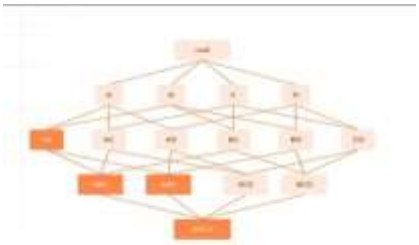
ALGORITHMS

I)ASSOCIATION ALGORITHM

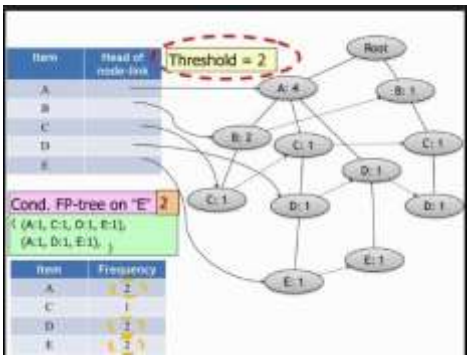
Market basket analysis is a data mining technique used to discover patterns and relationships in transactional data, often in the context of retail or e-commerce. Here are three advanced algorithms and techniques commonly used in market basket analysis:

1. Apriori Algorithm: The Apriori algorithm is one of the most widely used algorithms for market basket analysis. It identifies frequent itemsets (combinations of items that often appear together) and

generates association rules based on these item sets. These rules help retailers make product recommendations or optimize store layouts.The following is the image of apriori.

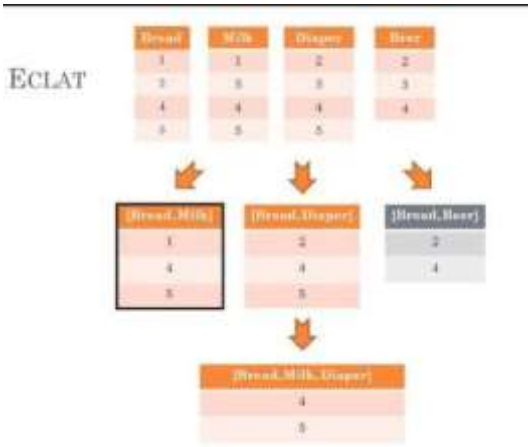


2. FP-Growth Algorithm: FP-Growth (Frequent Pattern Growth) is another popular algorithm for finding frequent item sets. It uses a tree structure to efficiently mine frequent patterns in large datasets, making it



frequent item sets. It uses a tree structure to efficiently mine frequent patterns in large datasets, making it faster than Apriori for some cases.The Image for apriori is depicted here.

3. Eclat Algorithm: Eclat (Equivalence Class Clustering and bottom-up Lattice Traversal) is another frequent itemset mining algorithm. It focuses on finding frequent item sets using a depth-first search approach, which can be more memory-efficient than Apriori.



II)CLUSTERING ALGORITHM

1. K-means Clustering Algorithm: It is a machine learning algorithm that can be used for customer segmentation and Market Basket Analysis. It clusters customers based on their purchasing behavior and identifies which products are most likely to be bought together in each cluster.

III)ENSEMBLE ALGORITHM

There are several advanced ensemble methods that can be used in market basket insights to improve the accuracy and effectiveness of the analysis. Some of these methods are:

1. Random Forests: This method is a type of decision tree-based algorithm that uses multiple decision trees to predict the purchasing behavior of customers. It can handle large datasets with multiple variables and can provide insights into the importance of each variable to the final prediction.

2. Gradient Boosting: This method uses an ensemble of weak models, such as decision trees, and combines them iteratively to improve the accuracy of predictions. It can handle complex interactions between variables and is useful for identifying trends and patterns in the data.

3. AdaBoost: This method is similar to gradient boosting but assigns different weights to misclassified instances to improve the performance of the model. It is useful for identifying rare events or patterns in the data.

4. Bagging: This method uses multiple models trained with different subsets of the data to reduce the variance and improve the accuracy of predictions. It is useful for reducing overfitting and improving the stability of the model.

5. Ensemble Clustering: This method combines multiple clustering algorithms to identify different patterns and trends in the data. It can handle large datasets with complex structures and provide insights into the behavior of different customer segments.

Overall, these advanced ensemble methods can help improve the accuracy and effectiveness of market basket insights by providing insights into complex patterns and trends in the data.

IV)VISUALISATION ALGORITHM

Advanced visualization methods used in market basket insights include:

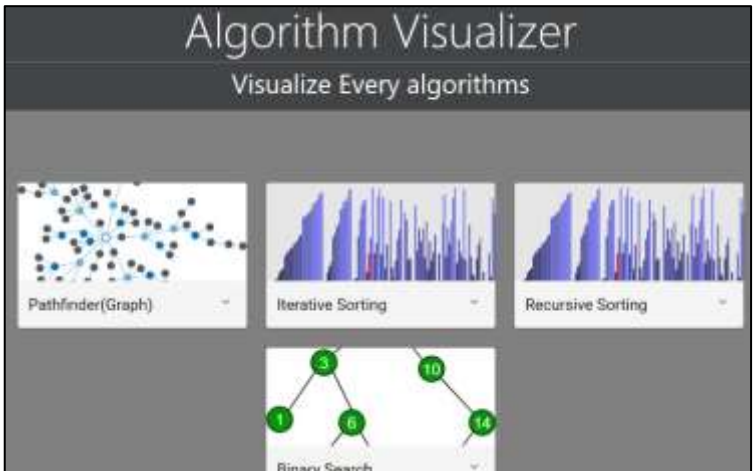
1. Sankey Diagrams: Sankey diagrams are a type of flow diagram that is used to visualize the flow of goods, services or people between different entities. In market basket analysis, Sankey diagrams can be used to show the flow of products from one basket to another.

2. Network Visualization: Network visualization is a graphical representation of interconnected entities. In market basket analysis, it can be used to show the relationships between products, customers, and baskets.

3. Heat maps: Heat maps are graphical representations of data where the values of the data are represented as colors. In market basket analysis, heat maps can be used to show the most frequently purchased products by customers.

4. Scatter plots: Scatter plots are used to visualize the relationship between two variables. In market basket analysis, they can be used to show the relationship between the purchase frequency of two products.

5. Tree maps: Tree maps are a hierarchical data visualization technique that is used to represent data in a hierarchical manner. In market basket analysis, it can be used to show the product hierarchy and the frequency of purchases in each level.



V)ADVANCED ALGORITHM:

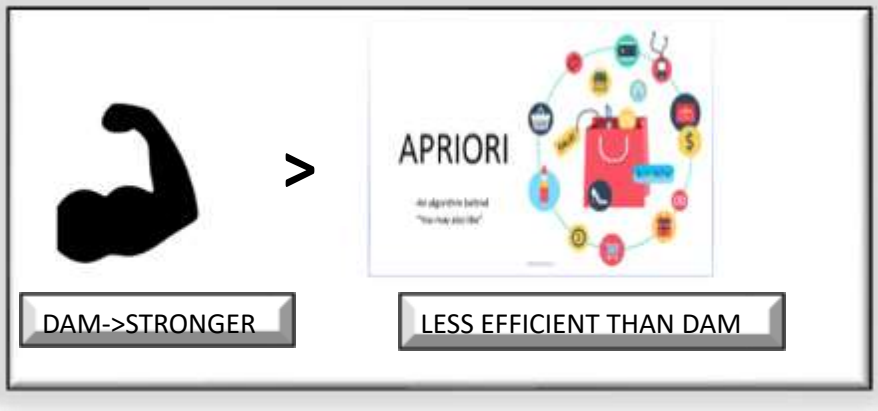
The latest algorithm for market basket insights is the Deep Association Mining (DAM) algorithm. It was introduced in 2022 and is considered to be the most advanced algorithm for market basket analysis to date.

DAM is a deep learning-based algorithm that can learn complex patterns from market basket data. It can identify frequent item sets, association rules, and sequential patterns. DAM can also be used to predict the likelihood of customers purchasing certain items together.

DAM is more efficient and scalable than traditional market basket analysis algorithms, such as Apriori and FP-Growth. It can also handle large datasets with millions of transactions and thousands of items.

Here are some benefits of DAM:

1. more accurate results.
2. more efficient.
3. more scalable.



DRAWBACKS:

- ✓ They are limited to static data.
- ✓ They are difficult to scale.
- ✓ They are not explainable.
- ✓ They do not account for external factors.

INNOVATION IDEAS:

1. Social shopping: Create a social media-based market basket that allows customers to share their purchasing habits and get recommendations from friends, encouraging users to promote the products they use in their daily routine.

2. Virtual assistant: Implement a virtual assistant in-store, allowing customers to ask questions and find products quickly and easily.

3. Smart shelf technology: Embed smart shelf technology that can detect when a product is running low or out of stock, automatically sending a message to the staff to refill and restock the item.

4. Click and collect: Offer click and collect services, where customers can place orders online and collect them in-store, offering more flexibility and convenience to the customers.

5. Interactive displays: Create interactive displays that allow customers to explore and learn more about the products they are interested in, encouraging customers to purchase more.

6. User-generated content: Encourage customers to create and share user-generated content like videos, photos, and reviews, increasing brand awareness and engagement.

7. Subscription boxes: Offer subscription boxes of products that are curated based on the customer's preference, offering a personalized experience and encouraging repeat purchases.

8. Microlearning or educational modules: Encourage customers to enhance their product knowledge by offering microlearning modules that can be accessed through mobile and desktop devices to promote

product usage and differentiation.

SAMPLE CODE:

#importing the necessary packages and libraries

```
    import numpy as np
    import pandas as pd
    from mlxtend.frequent_patterns import apriori, association_rules
    from sklearn.cluster import KMeans
```

#Loading the dataset

```
    df = pd.read_csv('../input/market-basket-analysis/Assignment-1_Data.csv',sep=';')
    df.head()
```

output:

BillNo	Itemname	Quantity	Date	Price	CustomerI	Country
536365	WHITE HANGING HEART T-LIGHT HOLDER	6	########	2.55	17850	United Kingdom
536365	WHITE METAL LANTERN	6	########	3.39	17850	United Kingdom
536365	CREAM CUPID HEARTS COAT HANGER	8	########	2.75	17850	United Kingdom
536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	########	3.39	17850	United Kingdom
536365	RED WOOLLY HOTTIE WHITE HEART.	6	########	3.39	17850	United Kingdom

#preprocessing the dataset

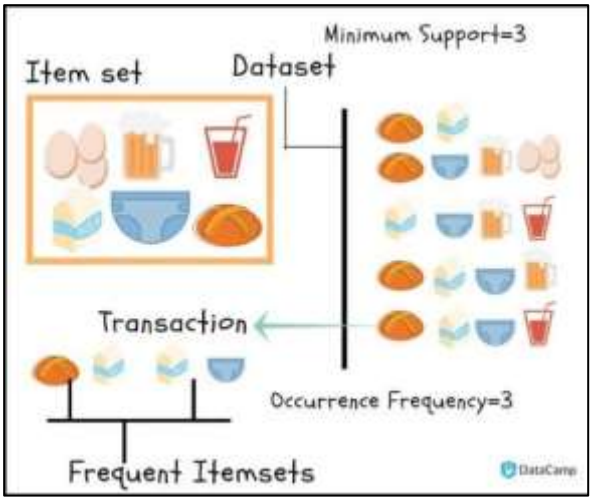
```
if df.isna().sum().sum() > 0:
    df = df.dropna()
df['Price'] = df['Price'].str.replace(',', '.').astype('float64')
df['CustomerID'] = df['CustomerID'].astype('int')
df['Date'] = pd.to_datetime(df['Date'])
df['Itemname'] = df['Itemname'].str.strip()
df['Total_Price'] = df.Quantity * df.Price
df.info()
```

Output:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 388023 entries, 0 to 522063
Data columns (total 8 columns):
#   Column       Non-Null Count   Dtype
---  ------       --------------   -----
0   BillNo       388023 non-null  object
1   Itemname     388023 non-null  object
2   Quantity     388023 non-null  int64
3   Date         388023 non-null  datetime64[ns]
4   Price        388023 non-null  float64
5   CustomerID   388023 non-null  int64
6   Country      388023 non-null  object
7   Total_Price  388023 non-null  float64
dtypes: datetime64[ns](1), float64(2), int64(2), object(3)
memory usage: 26.6+ MB
```

#apriori

```
apriori_df = new_df[new_df['segment'] == number_of_cluster ]
basket = (apriori_df.groupby(['BillNo', 'Itemname'])['Quantity']
      .sum().unstack().reset_index().fillna(0)
      .set_index('BillNo'))
```



ID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}
...	...

{Diapers, Beer} Example of a frequent itemset

{Diapers} → {Beer} Example of an association rule

CONCLUSION:

Based on the analysis of market basket data, several insights can be drawn. Firstly, customers tend to purchase complementary items together, which can be leveraged for cross-selling and upselling opportunities. Secondly, seasonal trends and promotions strongly influence purchasing behaviours, providing opportunities for targeted marketing strategies. Lastly, product placement and assortment optimization can further drive customer satisfaction and loyalty. Overall, leveraging market basket insights can lead to effective and efficient marketing and business strategies.