ARTIFICIAL INTELLIGENCE

PROJECT NAME:--MARKET BASKET INSIGH

PHASE 5 SUBMISSION BY:- JAGAN VIGNESH.B

NM ID:- DB66480CB493E7DA570430FCB88B1F03

YEAR:- 3RD

DEPT:- COMPUTER SCIENCE AND ENGINEERING (C.S.E)

"TABLE OF CONTENT"

A Market Basket Insights problem statement typically involves analyzing transaction data to gain a deeper understanding of customer purchasing behavior. The goal is to uncover patterns, associations, and trends in the products or services that customers buy together. This information can be valuable for various business objectives, such as improving marketing strategies, optimizing inventory management, and enhancing the overall customer experience.

Here's a general problem statement for Market Basket Insights:

01:- PROBLEM STATEMENT:

"In today's competitive market, businesses strive to understand their customers better and make data-driven decisions. One critical aspect of this is gaining insights into customer purchasing behavior. We aim to analyze transaction data to uncover meaningful associations and patterns in the products or services that customers purchase simultaneously. By doing so, we can address the following key objectives:

Cross-Selling and Upselling: Identify product pairs or groups that are frequently purchased together. This information can be used to create targeted cross-selling and upselling strategies, leading to increased revenue.

Inventory Management: Improve inventory management by identifying complementary products that should be stocked together. This can help reduce out-of-stock situations and minimize carrying costs.

Personalized Recommendations: Develop personalized product recommendations for customers based on their past purchase history and the purchasing behavior of similar customers.

Marketing Strategies: Optimize marketing campaigns by understanding which products tend to attract similar customer segments. This enables the creation of more effective and targeted marketing materials.

Customer Experience: Enhance the overall customer experience by offering relevant bundle deals, discounts, or promotions to customers based on their past buying habits.

The primary data source for this analysis will be transaction data, which includes details of purchases made by customers. This data will include information about the items bought, the date and time of purchase, and customer identifiers.

To address these objectives, we will employ various data mining and machine learning techniques, such as association rule mining, collaborative filtering, and clustering. The results of this analysis will provide actionable insights that can be used to drive business growth and enhance the customer experience.

The ultimate goal of this project is to enable our organization to leverage data-driven insights to make informed decisions that positively impact revenue, customer satisfaction, and operational efficiency."

The specific details of the problem statement may vary depending on the business, industry, and the available data. However, the general aim is to use transaction data to uncover valuable insights that can benefit the business in various ways, as outlined above.

02:-- DESIGN AND THINKING PROCESS

The design thinking process is a creative problem-solving framework used to develop innovative solutions by understanding the needs and desires of end-users. Applying the design thinking process to the context of Market Basket Insights involves a customer-centric approach to gain insights into customer purchasing behavior. Here's how you can use the design thinking process for Market Basket Insights:

1. **Empathize: Understand Customer Needs**

   - **User Research:** Start by collecting data on customer behavior. Analyze transaction data to understand what products or services customers buy and when.

   - **Customer Surveys and Feedback:** Gather information about customer preferences, shopping habits, and pain points through surveys and feedback mechanisms.

   - **User Personas:** Create personas to represent different customer segments based on their purchasing behavior. Understand their goals, motivations, and challenges.

2. **Define: Frame the Problem**

   - **Problem Statement:** Define the specific Market Basket Insights problem you want to solve, such as identifying product associations, optimizing inventory, or improving cross-selling.

   - **Key Metrics:** Determine the metrics and KPIs that will measure the success of your insights, e.g., revenue increase, inventory turnover, or customer satisfaction.

3. **Ideate: Generate Insights and Ideas**

   - **Brainstorming:** Gather a cross-functional team to brainstorm ideas for gaining insights from market basket analysis.

   - **Data Exploration:** Explore the transaction data to identify patterns, associations, and trends in customer purchasing behavior. Consider different data mining and machine learning techniques.

   - **Technology Considerations:** Explore the tools and technologies that can be used for data analysis and visualization.

4. **Prototype: Create Initial Insights**

   - **Data Analysis:** Use data mining techniques like association rule mining to discover product associations. Build initial insights from the data.

   - **Visualization:** Create visual representations (e.g., heatmaps, network graphs, or scatter plots) to present the discovered associations and patterns.

   - **User Testing:** Share the initial insights with a small group of stakeholders or end-users to gather feedback.

5. **Test: Refine and Validate Insights**

   - **Feedback Loop:** Incorporate feedback from user testing and iterate on the insights. Refine the associations and patterns.

   - **Statistical Significance:** Ensure that the discovered associations are statistically significant and not merely random occurrences.

   - **Cross-Validation:** Validate the insights on different subsets of the data to ensure their reliability.

6. **Implement: Operationalize Insights**

- **Integration:** Integrate the insights into your business processes, such as inventory management, marketing campaigns, or recommendations.

  - **Communication:** Share the insights and their implications with relevant teams and departments.

  - **Monitoring and Adaptation:** Continuously monitor the impact of implemented changes and adapt as needed.

7. **Feedback: Gather Ongoing Feedback**

  - **Customer Feedback:** Collect feedback from customers on the effectiveness of any changes made based on the insights.

  - **Data Iteration:** Continue to gather transaction data and update the insights as customer behavior evolves.

  - **Iterate and Improve:** Use feedback to iterate on your insights and improve their effectiveness.

Applying the design thinking process to Market Basket Insights ensures that your efforts are user-centric, data-driven, and focused on solving real customer needs. It encourages a cycle of continuous improvement, adapting to changing customer behavior and business requirements.

03:-- PHASES OF DEVELOPMENT**

The market basket insight phase of development typically refers to a stage in data analysis and business intelligence where you explore and analyze the items or products that customers tend to purchase together. This concept is particularly relevant in the context of retail and e-commerce businesses but can be applied in various industries to understand customer behavior and optimize business strategies.

Here are the key steps and elements involved in the market basket insight phase of development:

1. Data Collection: Gather data on customer transactions, including information about the items purchased, transaction timestamps, and customer demographics (if available).

2. Data Preprocessing: Clean and prepare the data for analysis. This may involve handling missing values, removing duplicates, and ensuring data consistency.

3. Association Rule Mining: Use techniques like Apriori or FP-growth to discover associations or relationships between items in the transaction data. These algorithms identify which items tend to be bought together.

4. Support, Confidence, and Lift: Evaluate the strength of the association rules using metrics like support, confidence, and lift. Support measures the frequency of itemsets in the data, confidence quantifies the likelihood of one item being purchased given the purchase of another, and lift measures how much more likely the two items are purchased together compared to what would be expected by chance.

5. Rule Selection: Choose relevant association rules based on predefined support, confidence, or lift thresholds. Filtering the rules helps focus on the most meaningful and actionable insights.

6. Visualization: Present the results through data visualization techniques such as scatter plots, heatmaps, or network diagrams to make the insights more understandable.

7. Interpretation: Interpret the discovered associations and patterns to gain insights into customer behavior. For example, you may discover that customers who buy pasta are also likely to buy pasta sauce.

8. Business Decision Making: Apply the insights to make informed business decisions. For instance, you can use this information to optimize product placement, develop targeted marketing strategies, create product bundles, or make pricing adjustments.

9. Continuous Improvement: The market basket insight phase is not a one-time effort. It should be an ongoing process where you regularly analyze new transaction data and adjust your strategies based on evolving customer preferences.

The market basket insight phase is part of a broader category of analysis known as market basket analysis or association analysis, and it is a valuable tool for improving sales, customer experience, and overall business performance.

04:-- DESCRIBE THE DATASET **

"Market basket insights" typically refer to the analysis of transaction data from retail or e-commerce businesses to understand customer purchase behavior. This type of analysis is commonly used to identify patterns and associations between products that are frequently purchased together, allowing businesses to make informed decisions about product placement, pricing, and marketing strategies. The specific dataset used for market basket analysis can vary widely, but here are some common attributes and details you might find in such a dataset:

1. Transaction Data: The dataset typically includes records of individual transactions. Each transaction represents a single purchase made by a customer and contains information such as transaction ID, purchase date, and customer ID (if available).

2. Product Information: Information about the products purchased in each transaction is included. This can include product names, SKU numbers, categories, descriptions, and prices. Each product in the dataset is usually associated with a unique identifier.

3. Customer Information (Optional): Some market basket analysis datasets may include information about the customers making the purchases, such as customer names, IDs, demographics, or loyalty program information.

4. Basket ID (Optional): In some datasets, a basket ID is used to group together products that were purchased in the same transaction. This is useful for identifying items that were purchased together in a single shopping trip.

5. Quantity Purchased: The quantity of each product purchased in a transaction is often recorded. This is essential for calculating metrics like item support and lift in association rule mining.

6. Time Stamp: A timestamp may be included to track the date and time of each transaction. This can be useful for analyzing purchase patterns over time, such as daily, weekly, or seasonal trends.

7. Additional Fields (Optional): Depending on the specific dataset and the goals of the analysis, additional fields may be included, such as discounts applied, payment methods, or store location.

The structure and format of market basket analysis datasets may vary depending on the source and purpose of the data. These datasets are typically used for techniques like association rule mining (e.g., Apriori algorithm) and collaborative filtering to discover relationships between products and make recommendations to customers.

05:-- DATA PREPROCESSING STEP**

Data Collection:

Gather transaction data from point-of-sale systems, e-commerce platforms, or other sources. Each transaction should contain information about the items purchased, the transaction ID, and the date and time of purchase.

Data Cleaning:

Clean the data to ensure its accuracy and completeness. This may involve handling missing values, dealing with duplicates, and correcting errors in the data.

Data Transformation:

Transform the data into a format suitable for market basket analysis. This typically involves organizing the data into a matrix or dataframe where each row represents a transaction and each column represents a product. The cells contain binary values indicating whether a product was in the transaction.

Item Encoding:

Encode the items in your dataset using a one-hot encoding scheme. This means representing each item as a binary (0 or 1) value in the matrix or dataframe. This step makes it easier to perform association rule mining.

Support and Confidence Thresholds:

Set minimum support and confidence thresholds. Support represents the proportion of transactions containing a specific itemset, while confidence measures the likelihood that an item Y is purchased when item X is purchased. These thresholds help filter out less relevant itemsets.

Apriori Algorithm:

Apply the Apriori algorithm or other association rule mining techniques to find frequent itemsets in your data. The Apriori algorithm is a common choice for this task.

Rule Generation:

Generate association rules based on the frequent itemsets. These rules indicate the likelihood of purchasing one item when another item is already in the basket.

Rule Evaluation:

Evaluate the generated rules based on various metrics such as lift, conviction, and lift-confidence ratio to determine the most meaningful and actionable rules.

Visualization:

Visualize the results using charts, graphs, or tables to make it easier to understand and interpret the findings. Common visualizations include support-confidence plots and network graphs.

Interpretation and Action:

Interpret the discovered rules and patterns to make informed business decisions. You can use these insights for inventory management, product placement, targeted marketing, and personalized recommendations.

Ongoing Analysis:

Continuously monitor and update your market basket analysis as new data becomes available. Market trends and customer preferences can change over time, so regularly re-evaluating the analysis is important.

Remember that market basket analysis is just one of many techniques for extracting insights from transactional data. The choice of algorithms, support, and confidence thresholds, and the interpretation of results should all be tailored to your specific business needs and goals.

06:-- ASSOCIATION ANALYSIS TECHNIQUES**

Transaction Data: Market basket analysis starts with transaction data, which typically includes a list of items purchased by customers in various transactions. Each transaction is represented as a set of items.

Support: Support is a measure that indicates how frequently an itemset (a collection of items) appears in the transactions. It is calculated as the number of transactions containing the itemset divided by the total number of transactions. High support values suggest that the itemset is commonly purchased.

Confidence: Confidence is a measure of how often a rule (an association between two itemsets) is true. It's calculated as the support of the combined itemset divided by the support of the antecedent itemset. High confidence indicates a strong association between the items.

Lift: Lift is a measure that quantifies how much more likely two items are purchased together than if they were purchased independently. A lift value greater than 1 indicates a positive association, while a value less than 1 suggests a negative association.

Apriori Algorithm: The Apriori algorithm is a widely used method for finding frequent itemsets and generating association rules. It employs a "bottom-up" approach, iteratively discovering frequent itemsets of increasing size based on user-defined support and confidence thresholds.

Pruning: Pruning is a technique used to reduce the number of itemsets and rules generated in order to focus on the most meaningful associations. One common form of pruning is setting minimum support and confidence thresholds.

Association Rules: Association rules are the outcome of market basket analysis and are typically written in the form "If {A} then {B}," where A and B represent itemsets. These rules reveal the relationships between items in the dataset.

For example, a simple association rule might be: "If a customer buys coffee and sugar, then they are likely to buy creamer." This kind of insight can be used to optimize product placement in a store, design targeted marketing campaigns, or create recommendations for online shoppers.

Market basket analysis can be a powerful tool for businesses to understand customer behavior, improve sales, and enhance the overall shopping experience. It has applications beyond retail, such as in healthcare for patient treatment recommendations, in web analytics for content recommendations, and in various domains where identifying associations in large datasets is valuable.

07,08:-- ASSOCIATION RULE AND BUSINESS IMPLICATIONS.**

Definition of Association Rules:

Association rules consist of three components:

Antecedent: The item or items that are purchased first or before the purchase of another item.

Consequent: The item that is purchased subsequently, often influenced by the antecedent item(s).

Support: A measure of how frequently a particular combination of items (antecedent and consequent) appears in the dataset.

Confidence: A measure of the likelihood that the purchase of the antecedent item(s) will result in the purchase of the consequent item.

Lift: A measure of how much more likely the consequent item is to be purchased when the antecedent item(s) are purchased compared to random chance.

Explaining Association Rules:

When explaining discovered association rules, you typically focus on the antecedent, consequent, support, confidence, and lift. Here's an example of how to explain a rule:

Rule: {Item A} -> {Item B}

Support: 5%

Confidence: 70%

Lift: 2.0

Explanation:

The antecedent, "Item A," is often purchased by customers.

When customers purchase "Item A," there is a 70% likelihood that they will also purchase "Item B."

This association between "Item A" and "Item B" is 2.0 times more likely than if these items were purchased randomly.

Business Implications:

Discovering and understanding association rules can have several important business implications:

a. Cross-selling and Upselling:

Businesses can use association rules to identify opportunities for cross-selling and upselling. For example, if customers often buy cereal and milk together, a retailer might place these items near each other to encourage additional purchases.

b. Inventory Management:

Retailers can optimize inventory management by ensuring that items commonly purchased together are stocked together. This reduces out-of-stock situations and improves customer satisfaction.

c. Pricing Strategies:

Companies can adjust pricing strategies based on association rules. For instance, if a rule indicates that customers who buy a particular item are highly likely to purchase a related item, the company may consider offering discounts on the related item to increase sales.

d. Personalized Marketing:

Association rules can be used to tailor marketing campaigns to individual customers. Customers who frequently purchase certain items together can be targeted with personalized product recommendations.

e. Store Layout:

Retailers can design store layouts to encourage customers to follow association rules, making it more likely that they'll purchase related items together.

f. Customer Experience:

Understanding association rules can enhance the overall shopping experience by making it more convenient for customers to find and buy related items.

In summary, association rules discovered through market basket analysis provide valuable insights into customer behavior and purchasing patterns, which businesses can leverage to improve their operations, marketing strategies, and customer satisfaction.

09:-- CODE IN PYTHON**

Market basket analysis can be performed in Python using libraries like `pandas` and `mlxtend`. Here's a simple example of how to perform market basket analysis with code:

```python
# Import the necessary libraries

import pandas as pd

from mlxtend.frequent_patterns import apriori

from mlxtend.frequent_patterns import association_rules


# Create a sample transaction dataset

data = {

    'Transaction_ID': [1, 2, 3, 4, 5],

    'Items': [

        'milk, bread, eggs',

        'bread, eggs, butter',

        'milk, bread',
```

```
        'bread, butter',

        'milk, eggs'

    ]

}


# Create a DataFrame from the dataset

df = pd.DataFrame(data)


# Preprocess the data by one-hot encoding

df['Items'] = df['Items'].str.strip().str.split(',')

df_encoded = pd.get_dummies(df['Items'].apply(pd.Series).stack()).sum(level=0)

df_encoded = df_encoded.astype(bool).astype(int)


# Apply the Apriori algorithm to find frequent itemsets

frequent_itemsets = apriori(df_encoded, min_support=0.2, use_colnames=True)


# Generate association rules

rules = association_rules(frequent_itemsets, metric='lift', min_threshold=1.0)


# Display the association rules

print("Association Rules:")

print(rules)

```

In this code:

1. We import the necessary libraries, including `pandas` for data manipulation and `mlxtend` for the Apriori algorithm and association rule generation.

2. We create a sample transaction dataset, where each row represents a transaction with a list of items purchased.

3. We preprocess the data by one-hot encoding the items using `pd.get_dummies`.

4. We apply the Apriori algorithm to find frequent itemsets. `min_support` is set to 0.2, meaning that itemsets occurring in at least 20% of the transactions are considered frequent.

5. We generate association rules using the `association_rules` function, and we specify the metric as "lift" with a minimum threshold of 1.0.

6. Finally, we print and display the association rules.

You can adapt this code to your own dataset and modify parameters such as `min_support` and the metric used for generating rules based on your specific analysis needs.

--:THANK YOU:--