GRT INSTITUTE OF ENGINEERING AND TECHNOLOGY, TIRUTTANI - 631209

Approved by AICTE, New Delhi Affiliated to Anna University chennai

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PROJECT TITLE

Market Basket Insights

COLLEGE CODE:1103

Phase :3

VISHNU . A

3rd yr, 5th sem

Reg no.:110321104058

achuachuachu619@gmail.com

MARKET BASKET ANALYSIS:

Market Basket Analysis (MBA) is a valuable data mining technique that offers actionable insights for various industries, particularly in the retail sector. It helps businesses understand customer purchasing behavior and discover associations between products. By identifying which items are frequently purchased together, businesses can make informed decisions regarding product recommendations, store layouts, marketing strategies, and more

DATA PREPARATION:

Collect transaction data: You need a dataset containing information about customer transactions. Each row represents a transaction, and the columns represent items/products purchased.

DATA PREPROCESSING:

Encode the data: Transform the data into a suitable format for analysis, often using a one-hot encoding. This means creating binary variables for each product where 1 indicates the presence of the product in the transaction, and 0 indicates absence.

FREQUENT ITEMSET MINING:

Apriori Algorithm: The Apriori algorithm is commonly used for finding frequent itemsets. It works by generating candidate itemsets and pruning those that do not meet the specified minimum support threshold (a measure of how frequently an itemset appears in the dataset).

ASSOCIATION RULE GENERATION:

From the frequent itemsets, association rules are generated. These rules consist of an antecedent (items on the left) and a consequent (items on the right) with a certain confidence and support level.

RULE PRUNING AND EVALUATION:

Prune rules: Remove rules that do not meet certain criteria, such as minimum confidence and lift.

Evaluate rules: Assess the strength and significance of the remaining rules using measures like support, confidence, and lift.

3.1BEGIN BUILDING YOUR MARKET BASKET ANALYSIS PROJECT: LOADING A DATASET:

INSTALL KAGGLE API:

If you haven't already, you need to install the Kaggle API package. You can do this using pip:

```
pip install kaggle
```

GET KAGGLE API CREDENTIALS:

You need a Kaggle account and a Kaggle API token to access datasets. You can get your API token by following these steps:

Go to your Kaggle account settings on the Kaggle website.

Scroll down to the "API" section and click on "Create New API Token."

This will download a JSON file containing your Kaggle API credentials.

Upload Kaggle API Credentials:

You need to upload your Kaggle API credentials JSON file to your working directory.

Now, you can use the Kaggle API to download the dataset from Kaggle and load it into your project.

How to load a dataset from Kaggle into your Python environment and begin building your

Market Basket Analysis project. To load a dataset from Kaggle, you typically use the Kaggle

API. Here's a step-by-step guide on how to do this:

PYTHON CODE FOR LOAD THE DATA:

```
import pandas as pd
from kaggle.api.kaggle_api_extended import KaggleApi
from zipfile import ZipFile

# Set your Kaggle API credentials
api = KaggleApi()
api.authenticate(api_key="your_api_key_here")  # Replace with your API key

# Download the dataset
dataset_name = "aslanahmedov/market-basket-analysis"
api.dataset_download_files(dataset_name)

# Unzip the downloaded file
with ZipFile(f"{dataset_name}.zip", "r") as zip_ref:
    zip_ref.extractall("data")  # Extract the contents to a "data" directory

# Load the dataset into a DataFrame
df = pd.read_csv("data/your_dataset_filename.csv")  # Replace with the actual CSV file name
```

Make sure to replace "your_api_key_here" with your actual Kaggle API key and "your_dataset_filename.csv" with the appropriate CSV file from the dataset.

Now you have loaded the dataset, and you can proceed with the steps mentioned earlier to perform Market Basket Analysis on it.

3.2PREPROCESSING A DATASET:

LOAD THE DATASET:

As mentioned earlier, you can load the dataset from Kaggle using the Kaggle API or by manually downloading it and loading it into your project.

EXPLORE THE DATA:

Start by examining the dataset to understand its structure, the columns it contains, and the nature of the data.

HANDLE MISSING VALUES:

Check for missing values in the dataset. Depending on the dataset and the library you're using, you can either drop rows with missing values or impute them with suitable values.

DATA ENCODING:

Market Basket Analysis typically requires transaction data in a specific format, often referred to as "basket data." Each row represents a transaction, and the columns represent items. Encode the data into this format, where each cell contains a binary value (1 if the item is present in the transaction, 0 if it's not).

DATA TRANSFORMATION:

Perform any additional data transformations as needed. This could include aggregating data, filtering out low-support items, or adjusting the dataset's structure to meet the Apriori algorithm's requirements.

DATA MINING:

Apply the Apriori algorithm or other frequent itemset mining techniques to find frequent itemsets in your dataset. These frequent itemsets are the foundation for generating association rules.

GENERATE ASSOCIATION RULES:

From the frequent itemsets, generate association rules using metrics such as support, confidence, and lift.

Python code snippet for preprocessing a Market Basket Analysis dataset:

```
import pandas as pd

from mlxtend.frequent_patterns import apriori

from mlxtend.frequent_patterns import association_rules

# Load the dataset

df = pd.read_csv("your_dataset.csv")

# Perform one-hot encoding

basket = pd.get_dummies(df, columns=["item_column"], prefix="", prefix_sep="")

# Group by transaction and sum the one-hot encoded items

basket = basket.groupby("transaction_id").sum()

# Convert item counts to 1 or 0

basket[basket >= 1] = 1

# Find frequent itemsets

frequent_itemsets = apriori(basket, min_support=0.1, use_colnames=True)

# Generate association rules

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

# Print the rules

print(rules)
```

Replace "your_dataset.csv" with the name of your dataset file and adjust the column names accordingly.

The code above demonstrates the preprocessing and analysis of a Market Basket Analysis dataset using the Apriori algorithm. Depending on your dataset's specifics, you may need to adapt these steps to your data.

3.3ANALYSIS PROCESS:

FREQUENT ITEMSET ANALYSIS:

Find frequent itemsets in the data, i.e., sets of items that are frequently purchased together.

Determine support values to identify how frequently these itemsets occur in the dataset.

ASSOCIATION RULE ANALYSIS:

Generate association rules that reveal the relationships between items.

Use metrics like confidence and lift to assess the strength and significance of these associations.

ITEM SET SIZE ANALYSIS:

Analyze the size (number of items) of frequent itemsets. Are customers more likely to buy small sets of items, or do they purchase larger sets?

SUPPORT ANALYSIS:

Explore the distribution of support values for different items or itemsets.

Identify items with high support to understand popular products.

CONFIDENCE ANALYSIS:

Examine confidence values for association rules.

Identify strong associations and potentially suggest cross-selling opportunities.

LIFT ANALYSIS:

Investigate lift values for association rules.

Discover which item combinations are truly associated and which ones are just random occurrences.

VISUALIZATIONS:

Create visualizations like heatmaps or network graphs to represent itemset associations, support, and confidence values.

RECOMMENDATION SYSTEMS:

Implement a basic recommendation system using association rules to suggest related products to customers during their shopping experience.

MARKET BASKET SEGMENTATION:

Segment customers based on their purchasing patterns, and analyze the frequent itemsets and association rules for each segment.

MARKET BASKET TIME ANALYSIS:

Analyze if there are temporal patterns in shopping behavior. Do item associations change during different times of the year or specific days of the week?

IMPACT OF PROMOTIONS:

Investigate how promotions or discounts affect purchasing behavior. Are there specific item associations that become more prevalent during promotional periods?

ITEM CO-OCCURRENCE ANALYSIS:

Explore which items tend to co-occur in the same transaction.

This can help with store layout optimization and product placement.

CROSS-SELLING STRATEGIES:

Develop strategies to cross-sell items based on the association rules. For example, if a customer adds bread to their cart, suggest butter or jam.

MARKET BASKET DIVERSITY ANALYSIS:

Analyze how diverse or homogeneous the items in a customer's basket are. Do customers tend to buy a variety of items, or do they stick to specific categories?

CUSTOMER SEGMENTATION:

Segment customers based on their purchasing behavior and analyze the frequent itemsets and association rules for each segment.

To perform these analyses, you'll need to adapt your code and data manipulation based on the specific questions and insights you're looking to extract from the dataset. Additionally, you can utilize data visualization tools and techniques to present your findings effectively.