

## PHASE 4 ASSIGNMENT

**PROJECT TITLE:** Feature selection, Model training, Evaluation of an dataset.

**PROBLEM DEFINITION:** Market basket insights, or market basket analysis, is the process of discovering associations and patterns in customer transaction data. The primary goal is to uncover relationships between products or items that are frequently purchased together. This project involves data preprocessing, feature engineering, model selection, training, and evaluation.

**GITHUB LINK:** <https://github.com/swetha5611/market-basket-insights>  
<https://github.com/swetha5611/innovation>

**DOCUMENT:** Building the project by Feature selection, Model training, Evaluation of an dataset.

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

Creating market basket insights involves understanding patterns in customer purchase behavior, and it typically requires a combination of feature selection, model training, and evaluation. Here's a step-by-step guide on how to approach this task:

### 1. Data Collection and Preprocessing:

- Gather the transaction data, which usually includes a list of items bought by each customer in different shopping baskets.
- Preprocess the data by handling missing values, removing duplicates, and ensuring data quality.

### 2. Feature Selection:

- Choose or engineer features that can help in understanding market basket insights. Common features may include:
  - Customer demographics (age, gender, location).
  - Product attributes (category, price, brand).
  - Time-related features (day of the week, time of day).
  - Transaction features (total spend, frequency of purchase).

### 3. Market Basket Analysis Techniques:

- Market basket analysis typically relies on techniques like Apriori, FP-growth, or association rule mining to discover patterns of items that are frequently purchased together. These techniques are essential for understanding the relationships between products.

### 4. Model Training:

- Apply the selected market basket analysis technique to the preprocessed data.
- Train the model to discover frequent itemsets, association rules, or patterns in the dataset.
- Set the appropriate parameters, such as minimum support and confidence thresholds, to control the sensitivity of the analysis.

### 5. Visualization:

- Create visualizations to represent the discovered patterns. Tools like heatmaps, network graphs, or simple bar charts can be helpful in understanding the relationships between items and customers.

### 6. Evaluation:

- Evaluate the model's performance using metrics like:
  - Support: The proportion of transactions containing a specific itemset.
  - Confidence: The probability that if a customer buys one product, they will also buy another.
  - Lift: The measure of how much more likely item B is purchased when item A is purchased compared to when item B is purchased independently of item A.
- Evaluate how well the patterns and rules generated from the model align with the actual behavior of customers.

### 7. Business Insights:

- Translate the results into actionable business insights. What can the market basket analysis reveal about customer behavior and preferences? How can this information be used to improve marketing, product placement, or inventory management?

### 8. Implementation:

- Implement strategies based on the insights gained. For example, you might adjust product placement in stores, create targeted marketing campaigns, or optimize inventory based on the market basket insights.

### 9. Continuous Monitoring:

- Market basket insights can change over time. Continuously monitor customer behavior and re-run the analysis to stay up-to-date with evolving trends and customer preferences.

## 10. Product Recommendation with Feature Selection, Model Training, and Evaluation:

```
# Import necessary libraries for recommendation
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report

# Load and preprocess your dataset, one-hot encode as described earlier

# Define features (X) and target (y)
X = df_encoded.drop('Target_Product', axis=1)
y = df_encoded['Target_Product']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                                    random_state=42)

# Train a recommendation model (e.g., a RandomForestClassifier)
model = RandomForestClassifier(n_estimators=100, random_state=0)
model.fit(X_train, y_train)

# Make predictions on the test data
y_pred = model.predict(X_test)

# Evaluate the recommendation model
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy}")

# You can also analyze the classification report for more detailed evaluation
print(classification_report(y_test, y_pred))
```

## 11. Communication:

Present your findings and recommendations to relevant stakeholders within your organization to ensure that the insights are put into practice.

**SUBMITTED BY,**  
**STUDENT REG NO: 711221104056**  
**NAAN MUDHALVAN: au711221104056**

