**Project 10 : Market Basket Analysis**

Market basket analysis using Apriori algorithm

**Introduction**

Market Basket Analysis is a data mining technique used to uncover patterns and associations between products in customer transactions. This project aims to apply Market Basket Analysis to understand customer purchasing behavior and identify cross-selling opportunities for a retail business.

Market basket analysis is a data mining technique that analyzes patterns of co-occurrence and determines the strength of the link between products purchased together. We also refer to it as frequent itemset mining or association analysis. It leverages these patterns recognized in any retail setting to understand the behavior of the customer by identifying the relationships between the items bought by them. To put it simply, market basket analysis helps the retailers know about the products frequently bought together so as to keep those items always available in their inventory.

The source from which these patterns are found is the vast amount of data [that is continually collected](https://www.turing.com/kb/how-data-collection-and-data-preprocessing-in-python-help-in-machine-learning) and stored. With frequent mining of the item set, it becomes easy to discover the correlation between items in huge relational or transactional datasets. It considerably helps in decision-making processes related to cross-marketing, catalog design, and consumer shopping analytics.

**Types of market basket analysis**

**1. Descriptive market basket analysis**

This type of market basket analysis offers actionable insights based on historical data. It is a frequently used approach that does not make any predictions but rates the association using statistical techniques between the products. We also refer to it as unsupervised learning based on the way it is modeled.

**2. Predictive market basket analysis**

Although “predict” and “analysis” make up the word predictive analysis, it actually works in reverse. It first analyzes and then predicts what the future holds. This type utilizes supervised learning models like [regression](https://www.turing.com/kb/regression-analysis-techniques-in-data-science) and classification. It is a valuable tool for marketers even if it is less used than descriptive market basket analysis.

So when we talk about the predictive market basket analysis, it considers items purchased in sequence to evaluate cross-sell. For instance, when a consumer purchases a laptop, they are more likely to buy an extended warranty with it.

**3. Differential market basket analysis**

Differential market basket analysis is a great tool for the competitive analysis that can help you determine why consumers prefer to purchase the same product from a particular platform even when they are labeled with the same price on both platforms.

This decision of the consumers is often based on several factors, as listed below.

* Delivery time
* User experience
* Purchase history between stores, seasons, time periods, and others.

Note: People without any expertise in data mining should consider confirming the results before sharing the results with stakeholders since it consists of various parameters and formulas that are taken into account at every step.

## **Terminologies used in market basket analysis:**

Here are some terminologies you should keep in mind while working with market basket analysis.

* **Itemset:** It refers to the set of items that are purchased together by a customer at the same time. By default, we state it as a logical rule with IF and THEN. For instance, IF (Bread, Butter), THEN (Milk). It is also possible that an item set may consist of no items that are usually ignored by all items in the dataset.
* **Support count:** It is the frequency of a particular item set appearing in the transaction database. It is also stated as a probability. For instance, if milk has a support count of 50 out of a possible 500 transactions, then the probability is 50/500 or 0.1.
* **Confidence:** It refers to the conditional probability that represents what items have a possibility of being purchased together. It finds application in the product placement strategy intending to increase profitability. For instance, by placing the high-margin items close to the related high-confidence items, retailers can increase the overall sales and revenue on purchases.
* **Antecedent:** The IF component written on the left-hand side or the item sets within the data are referred to as antecedents.
* **Consequent:** The THEN component or an item or itemset found in combination with the antecedent is called the consequent.

We can further understand antecedent and consequent with the below example.

Representation of antecedent and consequent.webp

## **Algorithms used in market basket analysis:**

Market basket analysis utilizes association rule **{IF} - > {THEN}** to predict the probability of certain products being purchased together. They count the item frequency occurring together and seek to find associations that occur more than expected.

Some algorithms that leverage these association rules are **AIS**, **Apriori**, and **SETM**.

Apriori is the commonly cited algorithm by the [**data scientist**](https://www.turing.com/jobs/remote-data-scientist) that identifies frequent items in the database. It is useful for unsupervised learning and requires no training and thus no predictions. This algorithm is used especially for large data sets where useful relationships among the items are to be determined.

You would be surprised to know that *Apriori algorithm leverages a shortcut namely Apriori property*. This shortcut states that all items in a frequent itemset must also be frequent. It helps in saving a lot of computational time.

The Apriori algorithm works in two steps that are illustrated below.

* It identifies the itemsets systematically that occur frequently in the dataset and support greater than the pre-specified threshold value.
* Next, it calculates the confidence of all possible rules. However, it only keeps those items states that have confidence greater than a pre-specified threshold.

It is further classified into three components.

* **Support**
* **Lift**
* **Confidence**

Let’s understand each one of them with an example of how to calculate market basket.

**Example:**

Let’s say there are 10 transactions for books and 8 transactions for pencils and 6 transactions are made for both products.

**1. Support**: It is the total number of transactions made for a particular product divided by the total number of transactions made. Zero represents no support while one represents the highest support. Higher the value of support, the greater the importance of the itemset in the data.

support(A⇒ B) =P(A ∪ B)

Support (Books) = Freq (Books)/Total transactions made

Support (Books) = 6/100 = 0.06%

**2. Confidence:** It is the ratio of combined transactions to individual transactions.

confidence(A⇒ B) =P(B|A)

Confidence (Books) = Combined transactions/Individual transaction

Confidence (Books) = 0.06/0.08 = 0.75

**3. Lift:** It is the ratio of the confidence percent to the support percent.

Lift = 0.75/0.10 = 7.5

* If the value of lift < 1, the combination is not bought by consumers frequently.
* If the value of lift >1, the combination is brought frequently by the consumers.
* If the value of lift = 1, then the purchase of antecedent makes no difference on the consequent.

Market basket analysis is used to search for the rules that result in a lift value greater than 1.

## **How does market basket analysis work?**

Market basket analysis is based on association rule mining which is

IF {}, THEN {} construct

It means that if a customer made a transaction that consisted of bread and butter, then they are likely to purchase milk too. However, before acting on any rule, the store manager or retailer must have sufficient evidence to back up the decision so that the results are beneficial. The above-discussed components namely support, confidence, and lift helps in measuring the strength of a rule to assist you in making an informed decision.

## **Market basket analysis example:**

Let’s understand the market basket analysis with an example of where the Apriori algorithm is implemented in the a rules package. It can be installed and run in **Python**.

**Import the necessary libraries**

import pandas as pd

import numpy as np

from mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

*#for viz*

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

*#to avoid warning*

import warnings

warnings.filterwarnings('ignore')

*#to display all feature if the number increase*

pd.set\_option('display.max\_columns', None)

# Data preprocessing

dataset = pd.read\_excel('/kaggle/input/market-basket-analysis/Assignment-1\_Data.xlsx')

Sample Output:

*# Get the top 10 item names by count*

top\_10\_items = data['Itemname'].value\_counts().nlargest(10).index

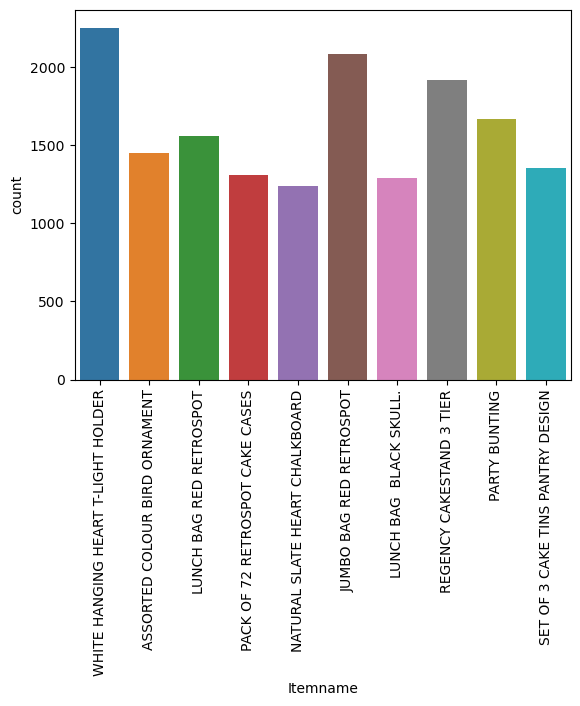
*# Filter the DataFrame to include only the top 10 item names*

df\_top\_10 = data[data['Itemname'].isin(top\_10\_items)]

*# Create a countplot for the top 10 item names*

ax=sns.countplot(data=df\_top\_10, x='Itemname')

plt.xticks(rotation=90)



## **Benefits of market basket analysis**

The never-ending list of impeccable benefits that market basket analysis has to offer is widely being leveraged by organizations around the world. This is also the reason one can notice a spike in the hiring of [ML engineers](https://www.turing.com/hire/ml-engineers) in companies around the world.

* **Personalized recommendations**
* **Promotions and campaigns**
* **Customer behavior analysis**
* **In-store operations optimizations**
* **New marketing tactics**

**Conclusion**

Cross-selling and upselling is the secret mantra of the retail industry that pushed the consumer to buy more. Organizations are using this technique wisely and making billions by playing with the mind of the customer.