Njala University

Module: Pattern Recognition

Project: Apriori Algorithm

Course: Master Computer Science

Presented By: Babah Saccoh & Oyah Mariama Kamara

Lecturer: Dr AJ Fofana

03 March 2025

Introduction

Overview

Market Basket Analysis (MBA) is a data mining technique used to uncover relationships between items purchased by customers. It is widely used in retail to identify patterns in customer purchasing behaviour, such as which items are frequently bought together. These insights can be used for:

•**Product Placement**: Placing frequently co-purchased items close to each other.
•**Cross-Selling**: Recommending complementary products to customers.
•**Promotions**: Designing bundle offers or discounts for frequently co-purchased items

This project uses the **Apriori Algorithm**, a popular algorithm for association rule mining, to analyse a transactional dataset and generate actionable insights.

Objectives

- **Preprocess the Dataset**: Convert the transactional data into a format suitable for the Apriori algorithm.
- **Find Frequent Itemsets**: Identify itemset that frequently appear together in transactions.
- **Generate Association Rules**: Discover rules that describe the likelihood of one item being purchased given that another item has been purchased.
- **Visualize Results**: Create visualizations to better understand the relationships between items.

System Component

Dataset:

The dataset used in this project is a **transactional dataset** where each row represents a transaction, and each column represents an item purchased in that transaction. The dataset contains the following items:

✦ **Items**: Bread, Milk, Eggs, Diaper, Beer, Coke, etc.

✦ **Transactions**: Multiple rows, each containing a list of items purchased together

Tools and Libraries:

✦ **Python**: Programming language used for implementation.

✦ **Pandas**: For data manipulation and analysis.

✦ **MLxtend**: For implementing the Apriori algorithm and generating association rules. ✦ **Matplotlib and Seaborn**: For data visualization.

Methodology

- **Load the Dataset :** The dataset is loaded into a pandas DataFrame for analysis.

Implementation of Apriori Algorithm

```
#Import required liberies
import pandas as pd
# Load the dataset
df = pd.read_csv('data.csv')
print(df.head())
```

```
   corned_b   peppers bourbon  cracker   chicken    apples      coke
0    olives   bourbon    coke   turkey  ice_crea       ham  baguette
1    hering  corned_b  olives      ham    turkey   bourbon   peppers
2  baguette  sardines  apples  peppers   avocado  ice_crea   bourbon
3  baguette      soda  hering  cracker  heineken   peppers    apples
4  baguette      soda  hering  cracker  heineken  corned_b       ham
```

Methodology

- **Preprocess the Dataset :** The dataset is converted into a list of transactions, and then into a one-hot encoded format using TransactionEncoder

```
from mlxtend.preprocessing import TransactionEncoder
transactions = df.apply(lambda x: x.dropna().tolist(), axis=1).tolist()
te = TransactionEncoder()
te_ary = te.fit(transactions).transform(transactions)
df_encoded = pd.DataFrame(te_ary, columns=te.columns_)
```

Methodology

- **Apply the Apriori Algorithm:** The Apriori algorithm is applied to find frequent itemsets with a minimum support threshold of 0.2.

```
from mlxtend.frequent_patterns import apriori
frequent_itemsets = apriori(df_encoded, min_support=0.2, use_colnames=True)
```

- **Generate Association Rules:** Association rules are generated using a minimum confidence threshold of 0.7.

```
from mlxtend.frequent_patterns import association_rules
rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.7)
```

Visualization code

- **Visualize Results** The results are visualized using bar plots, scatter plots, and heatmaps.

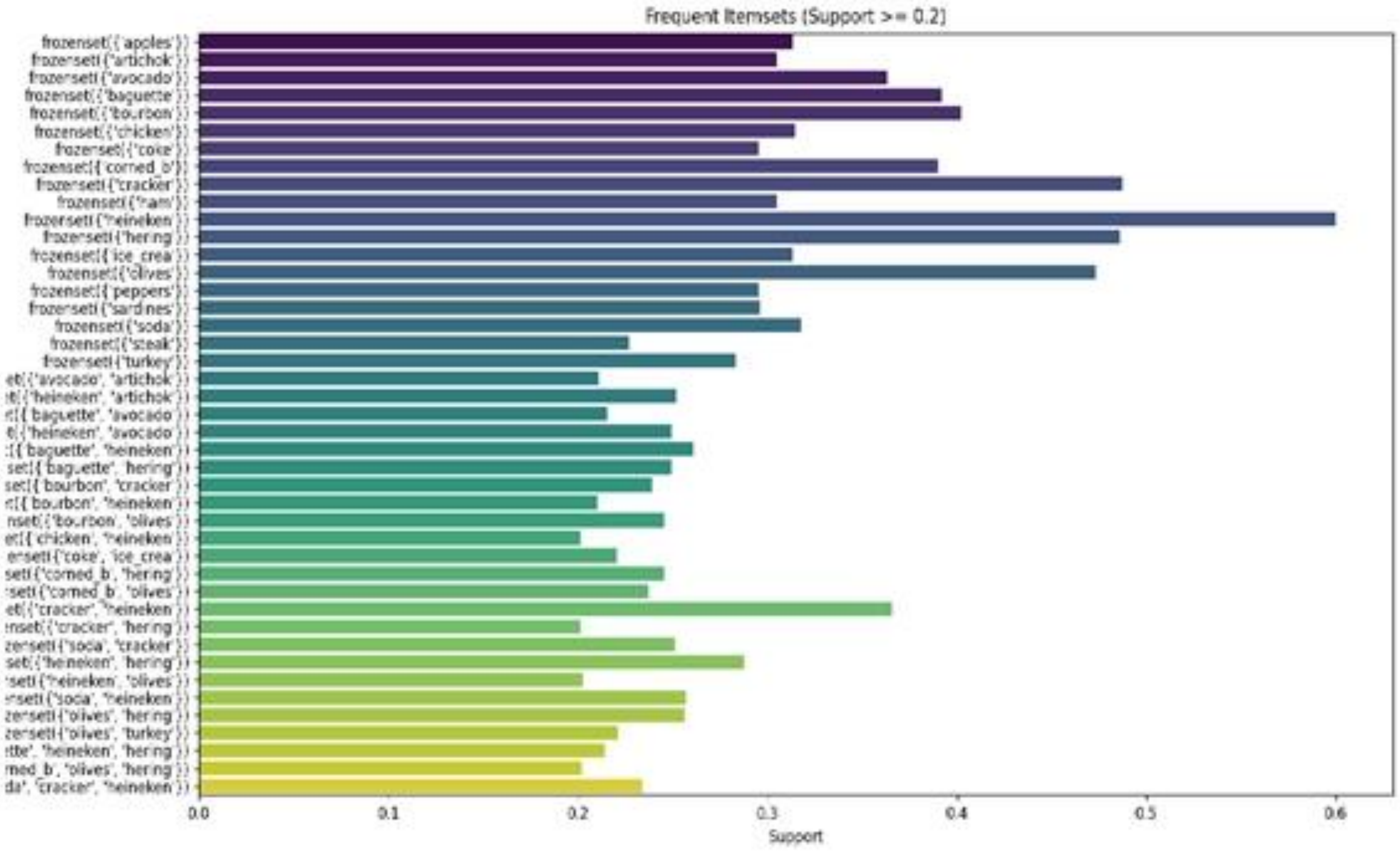
```
import matplotlib.pyplot as plt
import seaborn as sns

# Bar plot for frequent itemsets
plt.figure(figsize=(10, 6))
sns.barplot(x='support', y='itemsets', data=frequent_itemsets, hue='itemsets', palette='viridis', legend=False)
plt.title('Frequent Itemsets (Support >= 0.2)')
plt.xlabel('Support')
plt.ylabel('Itemsets')
plt.show()

# Scatter plot for association rules
plt.figure(figsize=(10, 6))
sns.scatterplot(x='support', y='confidence', size='lift', data=rules, hue='lift', palette='viridis', sizes=(20, 200))
plt.title('Association Rules (Support vs Confidence)')
plt.xlabel('Support')
plt.ylabel('Confidence')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.show()

# Heatmap for association rules
pivot_table = rules.pivot(index='antecedents', columns='consequents', values='confidence')
plt.figure(figsize=(10, 6))
sns.heatmap(pivot_table, annot=True, cmap='viridis', fmt='.2f')
plt.title('Association Rules Heatmap (Confidence)')
plt.xlabel('Consequents')
plt.ylabel('Antecedents')
plt.show()
```

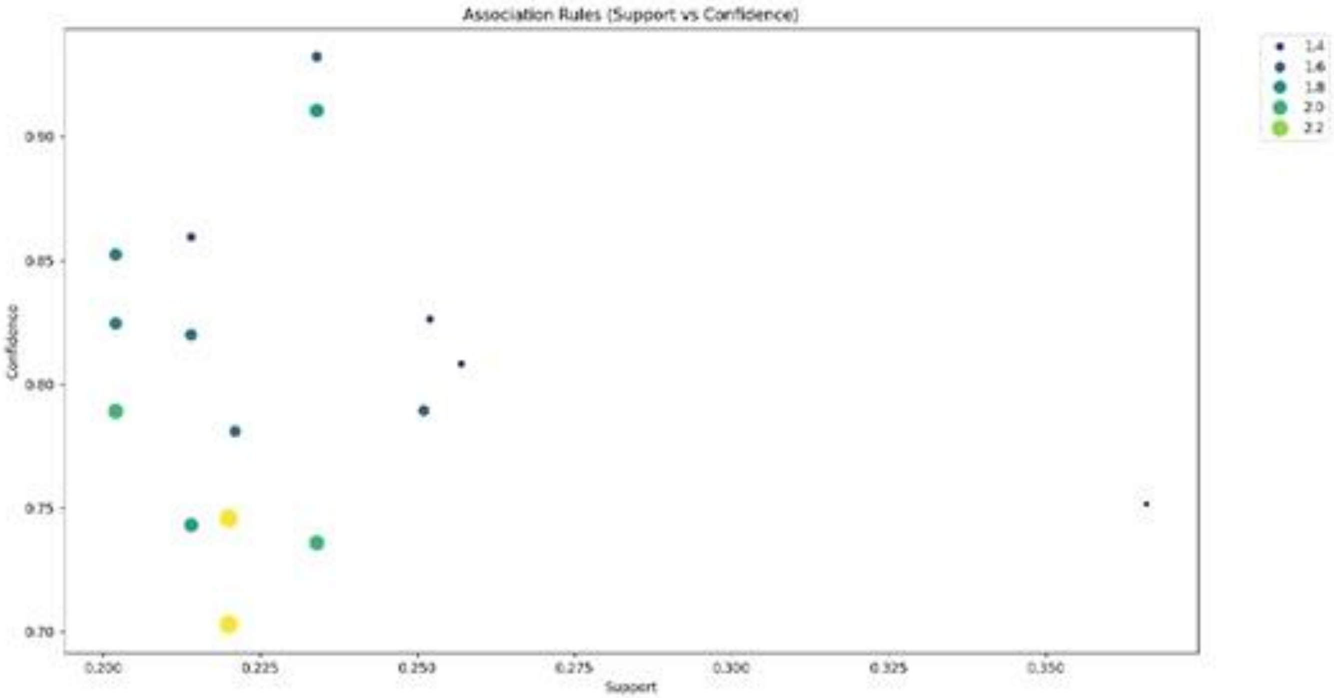
Frequent Items



The Apriori algorithm identified the following frequent items:

{Bread, Milk}: Support = 0.5
{Bread, Diaper}: Support = 0.5
{Milk, Diaper}: Support = 0.5
{Diaper, Beer}: Support = 0.5

Association Rules



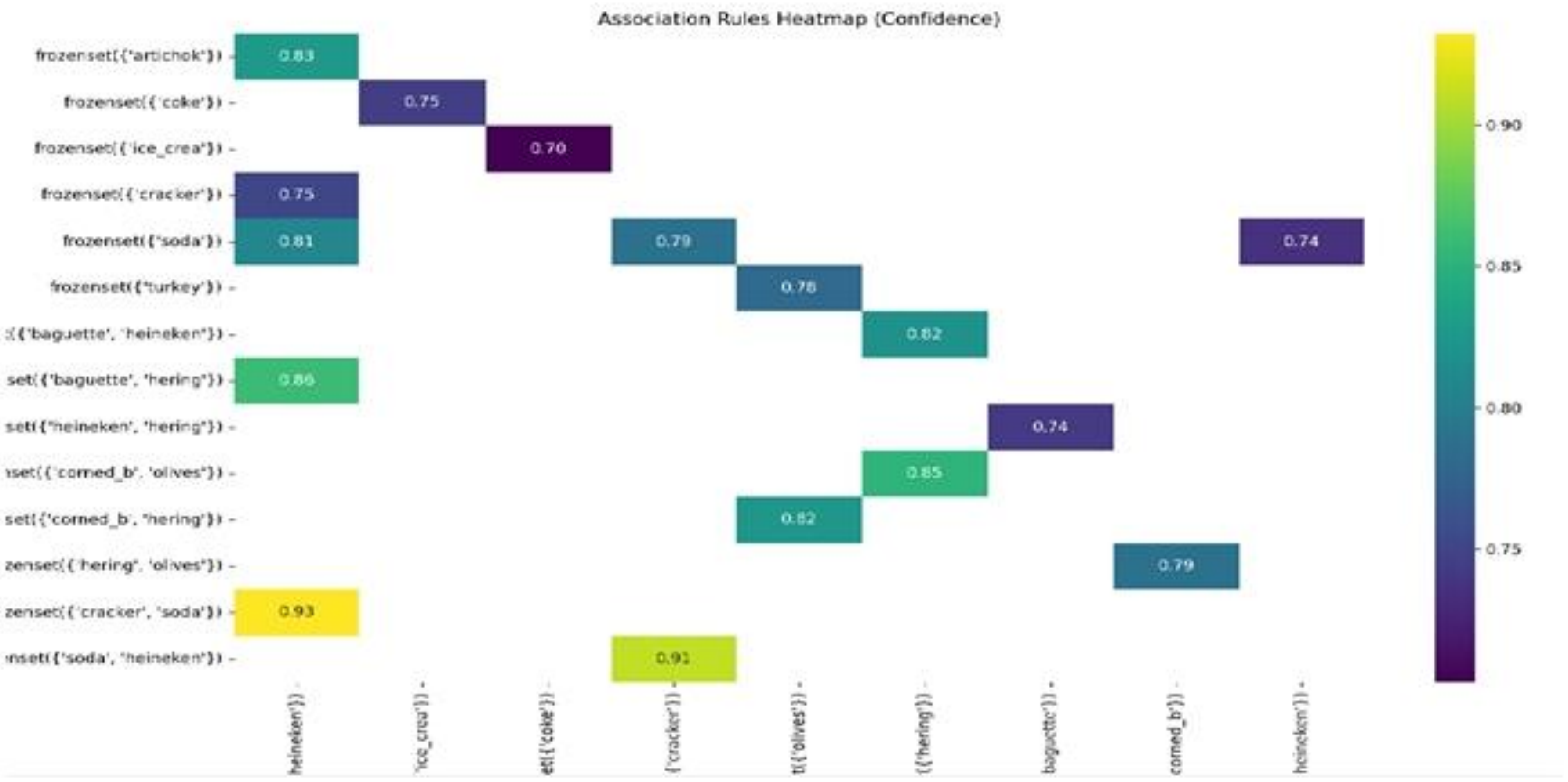
The following association rules were generated:

{Bread} -> {Milk}: Confidence = 0.67, Lift = 1.33

{Milk} -> {Bread}: Confidence = 0.67, Lift = 1.33

{Diaper} -> {Beer}: Confidence = 0.75, Lift = 1.5

Visualizations



Insights

Frequent Items:

✞ Customers frequently purchase {Bread, Milk} and {Diaper, Beer} together.

Association Rules:

✞ Customers who buy Bread are likely to buy Milk (67% confidence). ✞

Customers who buy Diaper are likely to buy Beer (75% confidence).

Recommendations:

✞ Place Bread and Milk close to each other in the store. ✞

Create bundle offers for Diaper and Beer.

Overview

This project successfully applied the Apriori algorithm to a transactional dataset to identify frequent items and generate association rules. The results provide actionable insights for optimizing product placement, cross-selling, and promotions. Future work could include:

•Experimenting with different support and confidence thresholds.

•Applying the algorithm to larger datasets.

•Integrating the results into a recommendation system.

Thank You

03 March 2025