

ASSIGNMENT NO. 01



Data Mining

Submitted To:

Dr. Ayyaz Hussain

Submitted By:

Iqra Fazilat (Sapid: 71036)

Semester:

Spring 2025

**Riphah International University Islamabad
Faculty of Computing**

Basket Analysis Report

Executive Summary

This report presents the results of a Basket Analysis conducted on a dataset of customer transactions. The analysis aimed to identify patterns and relationships between different products in the dataset using Classification-Based Association (CBA) rule mining. The results show that the model achieved an accuracy of **84.57%** and identified **292 association rules**, with **74 rules predicting the presence of chocolate** in a customer's basket.

Data Overview

The dataset used for this analysis consisted of **920 unique transactions**, with **16 products** in each transaction. The products included in the analysis were:

- - Apple
- - Butter
- - Corn
- - Eggs
- - Kidney Beans
- - Nutmeg
- - Sugar
- - Yogurt
- - Bread
- - Cheese
- - Dill
- - Ice cream
- - Milk
- - Onion
- - Unicorn
- - Chocolate

Model Performance

The model used for this analysis was a **Classification-Based Association (CBA) model**. The model achieved an accuracy of **0.8457**, indicating that it was able to correctly classify approximately **84.57%** of the instances in the dataset.

Confusion Matrix

The confusion matrix for the model is presented below:

	Predicted False	Predicted True
Actual False	488	21
Actual True	121	290

Classification Report

The classification report for the model is presented below:

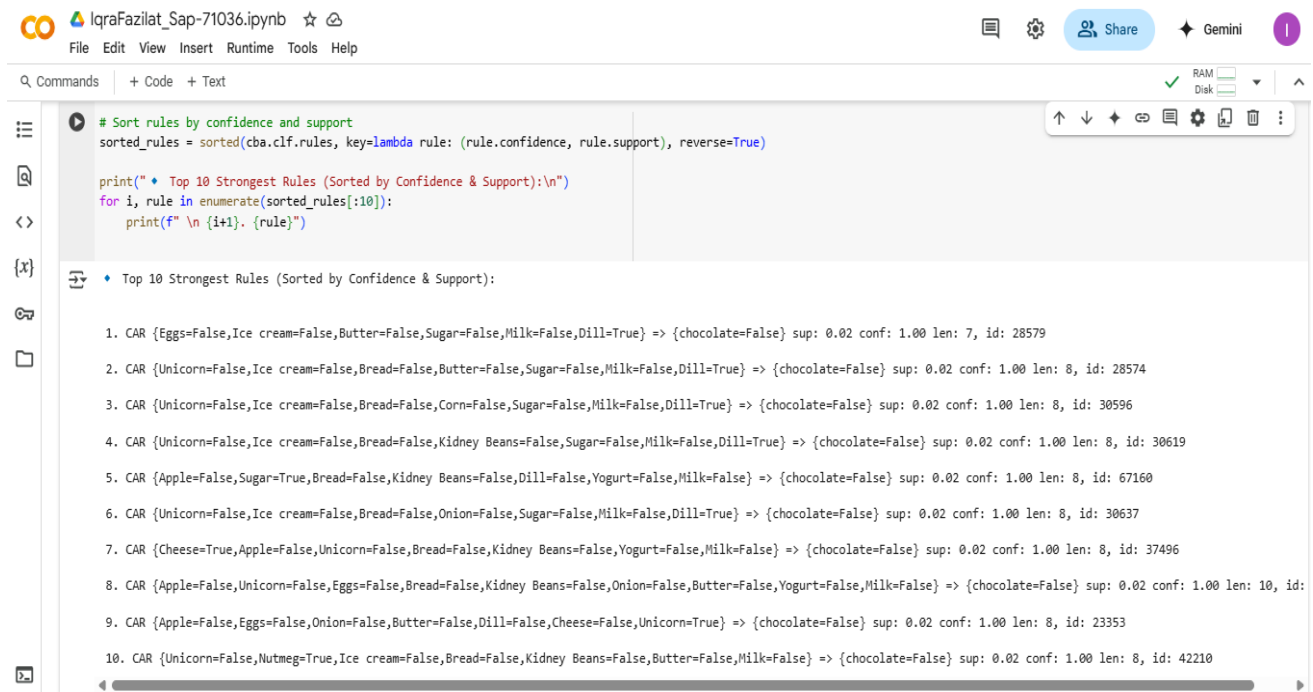
Class	Precision	Recall	F1-Score	Support
False	0.80	0.96	0.87	509
True	0.93	0.71	0.80	411
Accuracy	0.85			920
Macro Avg	0.87	0.83	0.84	920
Weighted Avg	0.86	0.85	0.84	920

Association Rules

The CBA model identified **292 association rules**, with **74 rules predicting the presence of chocolate** in a customer's basket. These rules provide insights into the relationships between different products in the dataset and can be leveraged for targeted marketing strategies.

Top 10 Association Rules

The top 10 association rules, sorted by confidence and support, are presented below:



The screenshot shows a Jupyter Notebook interface with a code cell and its output. The code cell contains the following Python code:

```
# Sort rules by confidence and support
sorted_rules = sorted(cba.clf.rules, key=lambda rule: (rule.confidence, rule.support), reverse=True)

print("\n • Top 10 Strongest Rules (Sorted by Confidence & Support):\n")
for i, rule in enumerate(sorted_rules[:10]):
    print(f"\n {i+1}. {rule}")
```

The output cell displays the top 10 association rules, each with its support, confidence, length, and ID. The rules are as follows:

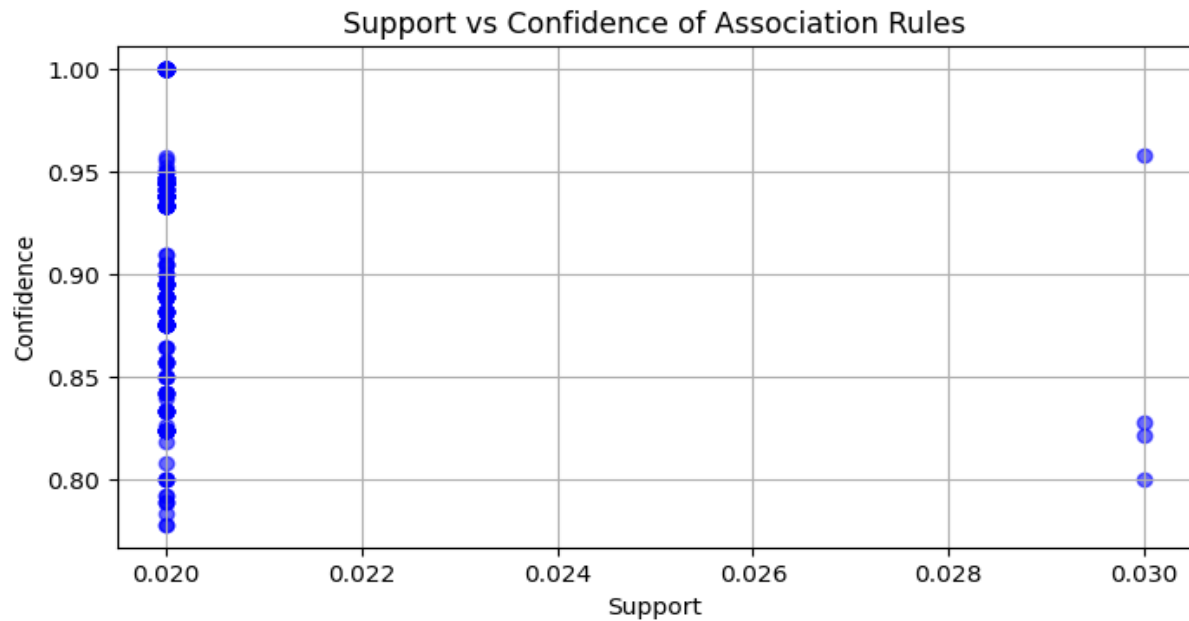
1. CAR {Eggs=False,Ice cream=False,Butter=False,Sugar=False,Milk=False,Dill=True} => {chocolate=False} sup: 0.02 conf: 1.00 len: 7, id: 28579
2. CAR {Unicorn=False,Ice cream=False,Bread=False,Butter=False,Sugar=False,Milk=False,Dill=True} => {chocolate=False} sup: 0.02 conf: 1.00 len: 8, id: 28574
3. CAR {Unicorn=False,Ice cream=False,Bread=False,Corn=False,Sugar=False,Milk=False,Dill=True} => {chocolate=False} sup: 0.02 conf: 1.00 len: 8, id: 30596
4. CAR {Unicorn=False,Ice cream=False,Bread=False,Kidney Beans=False,Sugar=False,Milk=False,Dill=True} => {chocolate=False} sup: 0.02 conf: 1.00 len: 8, id: 30619
5. CAR {Apple=False,Sugar=True,Bread=False,Kidney Beans=False,Dill=False,Yogurt=False,Milk=False} => {chocolate=False} sup: 0.02 conf: 1.00 len: 8, id: 67160
6. CAR {Unicorn=False,Ice cream=False,Bread=False,Onion=False,Sugar=False,Milk=False,Dill=True} => {chocolate=False} sup: 0.02 conf: 1.00 len: 8, id: 30637
7. CAR {Cheese=True,Apple=False,Unicorn=False,Bread=False,Kidney Beans=False,Yogurt=False,Milk=False} => {chocolate=False} sup: 0.02 conf: 1.00 len: 8, id: 37496
8. CAR {Apple=False,Unicorn=False,Eggs=False,Bread=False,Kidney Beans=False,Onion=False,Butter=False,Yogurt=False,Milk=False} => {chocolate=False} sup: 0.02 conf: 1.00 len: 10, id: 23353
9. CAR {Apple=False,Eggs=False,Onion=False,Butter=False,Dill=False,Cheese=False,Unicorn=True} => {chocolate=False} sup: 0.02 conf: 1.00 len: 8, id: 23353
10. CAR {Unicorn=False,Nutmeg=True,Ice cream=False,Bread=False,Kidney Beans=False,Butter=False,Milk=False} => {chocolate=False} sup: 0.02 conf: 1.00 len: 8, id: 42210

Visual Analysis

To further understand the patterns in the dataset, the following visualizations were created:

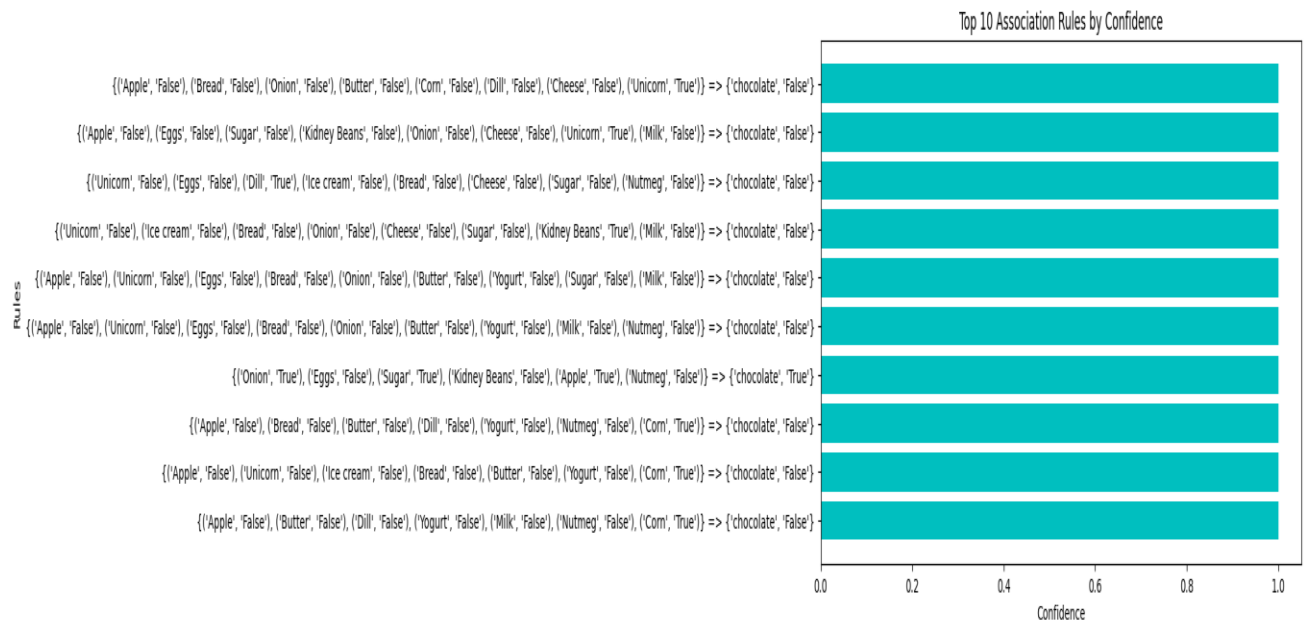
Scatter Plot of Support vs Confidence

The scatter plot below shows the relationship between support and confidence for the association rules, allowing us to identify the strongest and most reliable rules.



Bar Chart of Top 10 Association Rules

The bar chart below visualizes the top 10 association rules, sorted by confidence, providing a clear representation of the most influential rules in the dataset.



Conclusion

The basket analysis conducted on this dataset revealed significant patterns and relationships between different products. The model achieved an accuracy of **84.57%** and identified **292 association rules**, with **74 rules predicting the presence of chocolate** in a customer's basket. The insights gained from this analysis can be used to optimize marketing and sales strategies, improve product recommendations, and identify opportunities for **cross-selling and upselling**. Further refinements in the model and threshold tuning can lead to even more precise association rule generation.