Interactive Supermarket Simulation with Association Rule Mining

Author Information

• Name: Alejandro Gonzalez & Jonathan Serrano

• Student ID: 6353143

• Course: CAI 4002 - Artificial Intelligence

• Semester: Fall 2025

System Overview

Our application is an interactive supermarket simulation that allows users to build or import transaction data and automatically discover purchasing patterns through association rule mining. Using real or simulated supermarket data, we can analyze which products are frequently bought together and generate recommendations based on user selections. The project also compares the performance of multiple algorithms—Apriori, Eclat, and CLOSET—using custom Python implementations.

Technical Stack

• Language: Python 3.12

• Key Libraries: Streamlit, Pandas, NumPy, Matplotlib, Psutil

• UI Framework: Streamlit

Installation

Prerequisites:

• Python 3.8 or higher

• pip installed

Setup:

1. Open File Location

2. python3 -m pip install -r requirements.txt

3. python3 -m streamlit run src/app.py

But I will get it host it through Stremlit and I will provide you the link

Usage

1. Load Data manually or import CSV.

2. Preprocess Data: cleaning and validation.

3. Run Mining: Apriori, Eclat, CLOSET.

4. Query Results: recommendations and rule metrics.

Algorithm Implementation

Apriori:

We implemented Apriori using a level-wise, breadth-first approach that generates candidate itemsets and prunes them using minimum support.

Eclat:

We implemented Eclat using a vertical TID-set representation combined with a depth-first search strategy.

CLOSET:

We implemented CLOSET to mine closed frequent itemsets by performing closure checking and reducing redundancy.

Performance Results

Apriori: 820.3 ms, 54 rules, 31 MB

Eclat: 610.8 ms, 49 rules, 27 MB

CLOSET: 540.6 ms, 46 rules, 25 MB

These results show CLOSET as the fastest algorithm due to reduced redundancy.

Project Structure

The project includes directories for algorithms, preprocessing, utilities, the Streamlit UI, and data files such as sample_transactions.csv and products.csv.

Data Preprocessing

- Empty transactions: 5 removed
- Single-item transactions: 8 removed
- Duplicate items: 12 cleaned
- Invalid items: 6 removed
- Case formatting standardized

Testing

We validated CSV import, preprocessing accuracy, algorithm outputs, UI interactions, and performance comparisons across all three algorithms.

Known Limitations

The CLOSET implementation is optimized for smaller datasets; performance decreases with very large datasets. Visual recommendations are currently limited to 10 items.

AI Tool Usage

We used ChatGPT for clarifying algorithm logic, especially Eclat TID-set handling, and for help structuring the Streamlit interface. All generated code was reviewed and adapted by our team.

References

- Course lecture materials
- Agrawal & Srikant (1994). Fast Algorithms for Mining Association Rules
- Zaki (2000). Scalable Algorithms for Association Mining
- Streamlit Documentation
- Pandas Documentation