Midterm Project Report

Course: CS634 – Data Mining

Student Name: Naga Sathwik Sangaraju  
Student ID: NS2284  
  
1. Introduction

This project is a part of the CS634 – Data Mining course. It implements three algorithms Brute Force, Apriori, and FP-Growth to find frequent item sets and generate association rules from given datasets. The project demonstrates the use of these algorithms for pattern discovery and compares their efficiency in producing results.

# 2. Implementation Details

Key files used in this project include:

• io\_utils.py – Handles reading datasets, creating output folders, and user input.

• brute\_force.py – Implements the Brute Force algorithm for frequent itemset mining.

• apriori\_lib.py – Implements the Apriori algorithm using mlxtend library functions.

• fpgrowth\_lib.py – Implements the FP-Growth algorithm using mlxtend library functions.

• run\_all.py – The main file that allows dataset selection, input of minimum support and confidence, and runs all algorithms interactively.

# 3. Datasets

All datasets are stored in the 'data/' folder. Each dataset represents a collection of transactions containing sets of items. The available 5 datasets are amazon\_books.csv, homedepot.csv, mcdonalds.csv, netflix.csv, and starbucks.csv.

# 4. How to Run the Project

Steps to execute the project:

1. Open a terminal and go to the project directory:

cd path\to\cs634\_midterm\_Naga\_Sathwik\_sangaraju

2. Run the main script:

python src/run\_all.py

3. The program displays available datasets (1–5). Select one by entering the number, then input the minimum support and confidence values. All algorithms (Brute Force, Apriori, and FP-Growth) run automatically.

4. The results are printed in the terminal and saved in the 'outputs/' folder organized by dataset and algorithm.

# 5. Running with Jupyter Notebook

A Jupyter Notebook version ('datamining.ipynb') is available inside the 'notebooks/' folder. It can be opened using the command below:

jupyter notebook notebooks/datamining.ipynb

Run all cells in order. In the second cell, you can select the dataset and specify minimum support and confidence values. The notebook displays all results inline.

# 6. Converting Notebook to Python Script

We can convert the Jupyter Notebook into a Python script using nbconvert. Use the following command:

jupyter nbconvert --to script notebooks/datamining.ipynb

This command will generate a file named 'datamining.py' inside the same 'notebooks/' directory.

# 7. Results and File Structure

All generated files are saved in the 'outputs/' folder, grouped by dataset name and then by algorithm used. Each folder contains two CSV files: one for frequent itemsets and one for association rules. Example folder structure:

# cs634\_midterm\_Naga\_Sathwik\_Sangaraju │ ├── data │ ├── amazon\_books.csv │ ├── homedepot.csv │ ├── mcdonalds.csv │ ├── netflix.csv │ └── starbucks.csv │ ├── notebooks │ └── datamining.ipynb │ └── datamining.py ├── outputs │ └── (shows folders for each dataset and algorithm) │ ├── src │ ├── run\_all.py │ ├── io\_utils.py │ ├── brute\_force.py │ ├── apriori\_lib.py │ └── fpgrowth\_lib.py │ ├── requirements.txt ├── README.md └── CS634\_Midterm\_Report\_Naga\_Sathwik\_Sangaraju.docx Example 8. Conclusion

This project successfully demonstrates the implementation of Brute Force, Apriori, and FP-Growth algorithms to identify frequent itemsets and generate association rules. It also provides an interactive way for users to test different datasets and thresholds, helping them understand how these algorithms perform in practice.

# 9. Example output

Below is an example run on homedepot.csv with support=0.2 and confidence=0.4:

Available datasets:  
 1. amazon\_books.csv  
 2. homedepot.csv  
 3. mcdonalds.csv  
 4. netflix.csv  
 5. starbucks.csv  
Enter the number of the dataset you want: 2  
You selected: homedepot.csv  
Enter minimum support (0..1): 0.2  
Enter minimum confidence (0..1): 0.4  
  
=== [Brute Force] homedepot.csv ===  
Transactions: 25 | MinSupport=0.2, MinConfidence=0.4  
  
Frequent Itemsets (12 total):  
 {Hammer} support=0.6800  
 {Drill Machine} support=0.4400  
 {Nails Pack} support=0.3200  
 {Toolbox} support=0.3200  
 {Paint Brush} support=0.2800  
 {Safety Gloves} support=0.2000  
 {Screwdriver Set} support=0.2000  
 {Wrench Set} support=0.2000  
 {Hammer, Nails Pack} support=0.2800  
 {Hammer, Paint Brush} support=0.2800  
 {Hammer, Toolbox} support=0.2400  
 {Drill Machine, Safety Gloves} support=0.2000  
  
Association Rules (7 total):  
 {Paint Brush} -> {Hammer} support=0.2800 confidence=1.0000  
 {Safety Gloves} -> {Drill Machine} support=0.2000 confidence=1.0000  
 {Nails Pack} -> {Hammer} support=0.2800 confidence=0.8750  
 {Toolbox} -> {Hammer} support=0.2400 confidence=0.7500  
 {Drill Machine} -> {Safety Gloves} support=0.2000 confidence=0.4545  
 {Hammer} -> {Nails Pack} support=0.2800 confidence=0.4118  
 {Hammer} -> {Paint Brush} support=0.2800 confidence=0.4118  
  
=== [Apriori] homedepot.csv ===  
  
Frequent Itemsets (12 total):  
 {Drill Machine} support=0.4400  
 {Hammer} support=0.6800  
 {Nails Pack} support=0.3200  
 {Paint Brush} support=0.2800  
 {Safety Gloves} support=0.2000  
 {Screwdriver Set} support=0.2000  
 {Toolbox} support=0.3200  
 {Wrench Set} support=0.2000  
 {Drill Machine, Safety Gloves} support=0.2000  
 {Hammer, Nails Pack} support=0.2800  
 {Hammer, Paint Brush} support=0.2800  
 {Hammer, Toolbox} support=0.2400  
  
Association Rules (7 total):  
 {Safety Gloves} -> {Drill Machine} support=0.2000 confidence=1.0000 lift=2.2727  
 {Drill Machine} -> {Safety Gloves} support=0.2000 confidence=0.4545 lift=2.2727  
 {Nails Pack} -> {Hammer} support=0.2800 confidence=0.8750 lift=1.2868  
 {Hammer} -> {Nails Pack} support=0.2800 confidence=0.4118 lift=1.2868  
 {Hammer} -> {Paint Brush} support=0.2800 confidence=0.4118 lift=1.4706  
 {Paint Brush} -> {Hammer} support=0.2800 confidence=1.0000 lift=1.4706  
 {Toolbox} -> {Hammer} support=0.2400 confidence=0.7500 lift=1.1029  
  
=== [FP-Growth] homedepot.csv ===  
  
Frequent Itemsets (12 total):  
 {Hammer} support=0.6800  
 {Nails Pack} support=0.3200  
 {Drill Machine} support=0.4400  
 {Safety Gloves} support=0.2000  
 {Screwdriver Set} support=0.2000  
 {Paint Brush} support=0.2800  
 {Wrench Set} support=0.2000  
 {Toolbox} support=0.3200  
 {Hammer, Nails Pack} support=0.2800  
 {Drill Machine, Safety Gloves} support=0.2000  
 {Hammer, Paint Brush} support=0.2800  
 {Hammer, Toolbox} support=0.2400  
  
Association Rules (7 total):  
 {Nails Pack} -> {Hammer} support=0.2800 confidence=0.8750 lift=1.2868  
 {Hammer} -> {Nails Pack} support=0.2800 confidence=0.4118 lift=1.2868  
 {Safety Gloves} -> {Drill Machine} support=0.2000 confidence=1.0000 lift=2.2727  
 {Drill Machine} -> {Safety Gloves} support=0.2000 confidence=0.4545 lift=2.2727  
 {Hammer} -> {Paint Brush} support=0.2800 confidence=0.4118 lift=1.4706  
 {Paint Brush} -> {Hammer} support=0.2800 confidence=1.0000 lift=1.4706  
 {Toolbox} -> {Hammer} support=0.2400 confidence=0.7500 lift=1.1029  
  
  
  
  
  
  
  
10. Github repository

Link: https://github.com/ns2284/cs634\_midterm\_Naga\_Sathwik\_Sangaraju\_NS2284

Screenshots  
  
  
  
  
  
  
