

Assignment Instructions



Overview:

In this lab, you will explore association rule mining techniques using the Apriori and FP-Growth algorithms. You will analyze transactional data, identify frequent itemsets, and generate meaningful association rules. Additionally, you will use Seaborn to create visualizations that help interpret the mined patterns and insights. This lab will enhance your understanding of how association rule mining techniques are applied in real-world scenarios to uncover hidden patterns in data and present them effectively.

Lab Instructions:

Step 1: Data Preparation

- Create a new Jupyter Notebook and include a Markdown cell at the top with your name, course title, and lab assignment title.
- Select a publicly available transactional dataset such as:
 - Online Retail Dataset (UCI Machine Learning Repository)
 - Instacart Market Basket Dataset (Kaggle)
 - Retail Rocket Dataset (Kaggle)
 - Book-Crossing Dataset (Kaggle)
 - Movies and TV Show Purchases Dataset (Kaggle)
- Load the dataset and apply necessary data cleaning steps to ensure the data is ready for mining.
- Ensure that the dataset has item-level transaction records (e.g., product IDs or names) for compatibility with Apriori and FP-Growth algorithms.
- Create visualizations to explore the dataset and highlight important characteristics (e.g., Seaborn's barplot to show the most frequently occurring items in the dataset, Seaborn's heatmap to visualize item co-occurrence and highlight strong associations between items).

Step 2: Frequent Itemset Mining Using Apriori

- Use the Apriori algorithm to identify frequent itemsets in the dataset.
- Set appropriate support thresholds based on your dataset size and characteristics.
- Display the frequent itemsets with their corresponding support values.
- Use visualizations to present insights from the discovered itemsets (e.g., a Seaborn barplot to visualize the top N frequent itemsets with their support values).

Step 3: Frequent Itemset Mining Using FP-Growth

- Use the FP-Growth algorithm to identify frequent itemsets in the dataset.
- Apply the same support threshold as used in the Apriori algorithm.
- Compare the efficiency and output of the FP-Growth algorithm with Apriori.
- Use visualizations to present insights from the discovered itemsets (e.g., a Seaborn barplot to visualize the top N frequent itemsets with their support values).

Step 4: Generating and Analyzing Association Rules

- Using the frequent itemsets obtained from both algorithms:
 - Generate association rules using a defined confidence threshold.
 - Include association metrics such as *support*, *confidence*, and *lift*.
- Interpret the rules and provide insights about meaningful patterns.
- Use visualizations to illustrate key insights from the generated rules (e.g., Seaborn's scatter plot to visualize confidence vs. lift values to help identify high-confidence and strong association rules).

Step 5: Comparative Analysis

- Compare the results between Apriori and FP-Growth.
- Explain which algorithm was faster for your chosen dataset and why.
- Identify any challenges faced during the lab and how you resolved them.

Submission Instructions:

- Create a GitHub repository for this lab (e.g. MSCS_634_Lab_6).
- Add the following files to your repository:
 - Your Jupyter Notebook file (.ipynb).
 - A README.md file that briefly explains:
 - The purpose of your lab work.
 - Key insights from your analysis and association rules.
 - Any challenges faced or decisions made during the lab.
- Submit your GitHub repository link and ensure the repository is public or that you have granted instructor access.

Submission

T ▾ | ... | ¶ ▾ | ↶ ↷ ⊕ ▾

Drag and drop files here or click to add text

Word count: 0

Save and Close

Submit

