Assignment 1: Frequent Pattern Mining

Due Date: **October 24, 2023**

Instructions

- Submit your answer on Gradescope as a PDF file. Both typed and scanned handwritten answers are acceptable.

- Submit your answer on Gradescope as a PDF file. Both typed and scanned handwritten answers are acceptable.

- Late submissions will not be accepted. Exceptionally, each student may request a *one-day* extension for **one of the three** assignments, provided they contact the instructor and TA *before the deadline.*

- Cite all resources used. Plagiarism will be reported.

Problems

For this problem, consider the following set of transactions:

Transaction	Items
T1	{Python, Machine Learning}
T2	{Introduction to AI, JavaScript, C++}
T3	{Python, Machine Learning, Data Science, Introduction to AI, JavaScript}
T4	{Mathematics, Machine Learning}
T5	{Python, Data Science, Introduction to AI, Machine Learning}
T6	{Python, Data Science}
T7	{Mathematics, Machine Learning, Data Science}
T8	{Python, Mathematics}
T9	{Python, Machine Learning, Introduction to AI}
T10	{Introduction to AI, Mathematics, C++}

Problem 1: The Apriori Algorithm (50 points)

This problem can be done by hand or using a Python script (*hint: use the Python package **mlextend***).

1. **(20 points)** Given a minimum support of 3, apply the Apriori algorithm to the above transaction dataset to find frequent itemsets.

2. **(20 points)** Generate association rules based on the frequent itemsets with a minimum confidence level of 60 percent. (Hint: Confidence of a rule $A \rightarrow B$ is defined as the support count of $A \cup B$ divided by the support count of A.)

3. **(10 points)** Discuss the key strengths and limitations of the Apriori algorithm.

Problem 2: FP-Tree and FP-Growth (35 points)

Introduction to AI, Mathematics, C++ Use the below dataset to complete the rest of the assignment.

Transaction	Items
T1	{Python, Machine Learning, C++}
T2	{Introduction to AI, JavaScript, C++, Machine Learning}
T3	{Python, Data Science, Introduction to AI}
T4	{Mathematics, Machine Learning}
T5	{Python, Data Science, Introduction to AI, Machine Learning}
T6	{Python, Data Science}
T7	{Mathematics, Machine Learning, Data Science, JavaScript}
T8	{Python, Mathematics, Machine Learning}
T9	{Python, Machine Learning, Introduction to AI}
T10	{Python, Machine Learning, Introduction to AI}

1. **(15 points)** Construct the FP-Tree for the given transaction dataset. Describe the process and provide a visual representation of the final tree.

2. **(10 points)** Based on the FP-Tree, use the FP-Growth algorithm to generate the frequent itemsets. Please do NOT use a script for this.

3. **(10 points)** Compare the key strengths and limitations of the FP-Growth algorithm and the Apriori algorithm.

Problem 3: Constraint-based Frequent Pattern Mining (15 points)

If you wrote a script for Question 1, you are permitted to use it to answer this question. Consider a budget constraint of $3800 for purchasing books. Assume the cost of each item is as follows:

Item	Cost	Timeslot (50 mins)
Python	$200	8 am
Mathematics	$750	9 am
Machine Learning	$3000	1 pm
Introduction to AI	$1000	10 am
JavaScript	$400	8 am
C++	$500	8 am
Data Science	$500	9 am

Please use the dataset in Question 2 to answer this question.

1. **(5 points)** Generate frequent itemsets from the given dataset that comply with the budget and timeslot constraint using either the Apriori or FP-Growth algorithms. Assume that only classes that start at the same time introduce conflicts.

2. **(5 points)** Discuss whether the budget constraint mentioned above is anti-monotone or not. Justify your response and explain why it is useful to consider.

3. **(5 points)** Describe a constraint that could be applied to this dataset that will be either (a) anti-monotone if you answered monotone for the previous question, or (b) monotone if your previous answer was anti-monotone. Explain why this type of constraint is useful to consider.