Instructions:

1. **Write your name and roll number on the question paper.**
2. **Use of any electronic device or reading material is not allowed.**
3. **Write appropriate explanation/justification/steps wherever necessary.**
4. **Any form of plagiarism will be penalized.**
5. **If anything is not clear, or there is a mistake, or some question is incomplete, make appropriate assumptions, mention them and proceed.**

1. Create an arbitrarily random binary matrix of size 10×5, where each column denotes a document represented by a shingle. This matrix should contain around $30 - 50\%$ 1s and remaining 0s, spread in a random manner. Also create 4 arbitrary permutation vectors to permute the shingles. Using these, compute the (min-hash) signature matrix using the one-pass implementation algorithm, showing all the calculations and steps. Then compute the similarity of the first document with all the remaining documents using the two representations (shingles and signatures) and corresponding similarity measures as discussed in class. [21 **Marks**]

2. In a store, there are 10 items {'a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j'}. Create 10 random and distinct baskets using these items, with each basket containing 6–8 items. Execute the Apriori algorithm to find all frequent itemsets of size {1,2,3,4}, for support = 0.7. Using the frequent itemsets of size 2, create all possible association rules, and calculate their confidence and interestingness scores. [23 **Marks**]