**PSG COLLEGE OF TECHNOLOGY**

**DEPARTMENT OF APPLIED MATHEMATICS AND COMPUTATIONAL SCIENCES**

**M.Sc (TCS / SS) – 15XTO3- Data Science**

**PROBLEM SHEET-II – Finding similar items**

1. Find k shingles from the statement “A spokesperson for the Sudzo Corporation revealed today that studies have shown it is good for people to buy Sudzo products.” For k=3.
2. Find k shingles that contains only stop words for the statement in question 1.
3. Compute the Jaccard similarities of each pair of the following three sets:{1,2,3,4},{2,3,5,7}, and {2,4,6}.
4. From the characteristic matrix given below, find the signature similarity using hash functions h1(x) = x+1 mod 5 and h2(x) = 3x+1 mod 5.

