**Header:**

**MinHash Signature of Document:**

Pick p1 … pk permutations

Min[p1(D)], Min[p2(D)], etc

MinHash matrix is a matrix, where the columns are for each document, and the rows are for the permutations

IE. with document I and permutation j, the ith column and jth row = Min[pj(Di)]

Algorithm for Min-Hash Matrix:

Input: D1 … Dn

1. Pre-process and collect terms = { t1 … tm}
2. Randomly pick p1 … pk permutations on {1 … m}
3. For every Di:
   1. Compute Min[pj(Di)] for 1 =< j =< k