Lab Assignment 8 (Bloom Filter): CS2233

22nd October, 2025

Datasets: Please use the dataset in the attached txt file `dataset`. The dataset consists of streams of integers. You may consider all the numbers mentioned in the file as one stream. Size of universe $|\mathcal{U}| = 1000000$, $n = 800$.

- Implement `Bloom filter` with the following values of the sketch size $m = 2000, 3000, 4000$. Please use the appropriate values of the hash function as per the sketch size and number of items in the stream.

  There are two query files - `query_positive` consist of all those elements which are present in the datasets, and `query_negative` consist of all those elements which are not present in the datasets. The problem is to hash all elements in both query files and print the corresponding confusion matrices.