Jaccard score Vs percentage of data generated