**Logo, company name

Description automatically generated**

**COMP-4250 Big Data Analytics and Database Design**

**Project I**

**Report on: Mining Frequent Itemsets**

**Name:** Kenil Chovatiya **Student Id:** 105163073

Specification of my Computer:

* **Operating System:** Windows – 10
* **CPU Specs:** intel core i7 8th generation
* **RAM:** 16 GB

Explanation and Result:

There are several classes I created for developing entire project. I used JAVA to develop the code. Following are the classes and their description:

* **A\_Priori:** in this class I defined algorithm for A-Priori with all the phases of A-Priori to find the frequent pair.
* **PCY:** algorithm for PCY with all the phases including hash function to find frequent pairs in complete dataset.
* **Bucket:** used to hold the memory.
* **SetFile:** Generating hash function and implanting into dataset.
* **SetFile2:** To override hash function. (but never used)
* **Begin:** Used to generate value from dataset.
* **RunThis:** This is the main function of entire project. It calculates the time for both algorithm for different percentage of support threshold for given dataset and return frequent pair according to percentage of data set given to function. (Here, I choose 1%, 5% and 10 % of support threshold given data set of retail store.)

Here are the frequent pairs generated for chosen Support threshold from given dataset: \

Total number of buckets are 88161 for given dataset.

For 1% of support threshold frequent pairs are: (with number of repetition)

{(36, 38) =2790, (36, 39) =2037, (37, 38) =1046, (38, 39 )=10345, (38, 41) =3897, (39, 41) =11414, (38, 48) =7944, (39, 48) =29142, (32, 41) =3196, (39, 79) =1111, (36, 48) =1416, (41, 48) =9018, (48, 79) =893, (39, 89) =2749, (39, 101) =1400, (48, 89) =2798, (48, 101) =1311, (38, 110) =2725, (39, 110) =1759, (32, 39) =8455, (32, 48) =8034, (32, 38) =2833, (39, 65) =2787, (48, 65) =2529, (39, 60) =983, (41, 65) =995, (48, 110) =1380}

For 5% of support threshold frequent pairs are: (with number of repetition)

{(38, 39) =10345, (39, 41) =11414, (38, 48) =7944, (39, 48) =29142, (41, 48) =9018, (32, 39) =8455, (32, 48) =8034}

For 10% of support threshold frequent pairs are: (with number of repetition)

{(38, 39) =10345, (39, 41) =11414, (39, 48) =29142, (41, 48) =9018}

Time taken for A-Prior and PCY are in .xml file(attached). Also, it is listed below according to my Computer specification.

Graphical user interface, application

Description automatically generated