**Project Report**

To compile and execute follow the below steps:

Change the directory in cmd from home to src folder of in Project1 folder.

To Compile use the below command on command Line:

**Javac DPP.java**

To Execute use the below command:

**Java DPP P1InputData 3 4**

Pseudocode:

Start

Extract data from csv file and insert it into an array.

Insert the last column which contains class values in a separate array.

Create a method to find entropy and Information gain.

Using the above method find Information Gain for “k” the genes. Where “k” is top k genes of input data.

Sort the genes based on Information Gain decreasing order.

Split the genes into “m” intervals.

Itemize the final data.

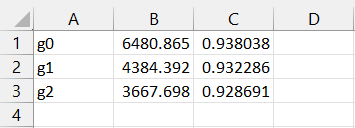
Write the output to respective csv and Text files.

Stop

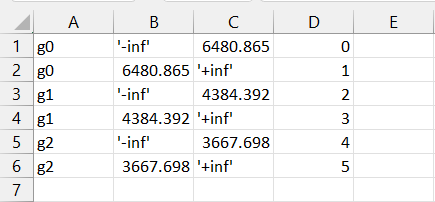
Outputs:

**Task1**

a) entropyRank.csv



b) entropyItemMap.csv



C) entropyItemizedData.txt

1,3,5,

1,3,5,

0,3,5,

0,3,5,

0,2,4,

0,2,4,

1,2,5,

0,2,4,

1,3,5,

0,3,4,

1,2,4,

1,3,5,

0,3,5,

0,3,5,

0,3,5,

0,2,4,

1,2,4,

0,2,4,

0,2,4,

0,2,4,

1,3,5,

1,3,5,

0,2,4,

0,2,4,

1,3,5,

0,2,4,

0,2,4,

1,3,5,

1,3,5,

1,3,5,

1,3,5,

0,2,4,

1,2,4,

1,3,5,

0,2,4,

0,2,4,

0,3,5,

0,2,4,

0,3,5,

1,3,5,

0,3,5,

0,3,5,

1,2,4,

1,2,4,

1,3,5,

1,2,4,

1,3,5,

0,2,4,

0,2,4,

1,3,5,

1,3,5,

1,3,5,

1,3,5,

0,3,5,

0,2,4,

0,2,4,

0,3,5,

0,2,5,

1,3,5,

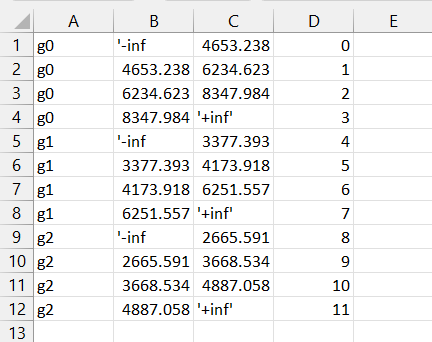
1,2,4,

0,2,4,

1,2,4,

**Task2**

a) equiDensityItemMap.csv



b) equiDensityItemizedData.txt

3,6,10,

3,7,10,

0,7,11,

2,7,11,

0,5,9,

0,4,8,

2,5,10,

0,4,9,

3,7,10,

1,6,9,

3,4,8,

3,7,11,

0,6,10,

1,6,10,

1,6,10,

1,5,8,

2,4,8,

1,4,8,

1,5,9,

0,4,8,

2,7,11,

2,7,11,

0,4,8,

0,4,8,

3,6,10,

1,5,9,

1,5,9,

3,7,11,

3,7,11,

3,7,11,

3,7,11,

1,4,8,

2,4,8,

2,7,11,

1,4,9,

0,4,9,

1,7,11,

0,4,8,

0,6,10,

3,6,10,

0,6,11,

1,6,10,

3,5,9,

3,4,9,

3,7,11,

3,5,9,

2,6,10,

2,5,9,

2,5,8,

2,6,10,

2,6,10,

3,7,11,

2,7,11,

0,6,10,

1,5,8,

0,4,8,

0,6,11,

1,5,9,

3,7,11,

2,5,8,

1,5,9,

2,5,9,