

Scalable Data Mining: Assignment 1

Pankaj Mishra, 17EC35034

Question 2

Here we use the concept of inverted indexing to make search faster. First part involves building a method to create an inverted indexed table for fast lookup. The second part is naive searching for number of records with a given attribute value. The final part is doing the same using Spark on the inverted indexed table, which is much faster in its performance.

NOTE: I was not able to run the first and third part of this question as the code threw a `OutOfMemoryError` everytime the program was executed. This happened with multiple approaches such as `groupBy`, `productElement`, etc. Therefore the output for these 2 parts has not been provided.