**BDA LAB 9**

**CE 142**

**Objective: Leveraging machine learning using Mahout like tools.**

**Introduction:**

**Apache Mahout, a scalable high performance machine learning framework**

Apache Mahout (mahout.apache.org) is a powerful and high performance machine learning framework for the implementation of machine learning algorithms. It is traditionally used to integrate supervised machine learning algorithms with the target value assigned to each input data set. Apache Mahout can be used for assorted research based applications including social media extraction and sentiment mining, user belief analytics, YouTube analytics and many related real-time applications.

In Apache Mahout, a ‘mahout’ refers to whatever drives or operates the elephant. The mahout acts as the master of the elephant in association with Apache Hadoop and is represented in the logo of the elephant. Apache Mahout runs with the base installation of Apache Hadoop, and then the machine learning algorithms are implemented with the features to develop and deploy scalable machine learning algorithms. The prime approaches, like recommender engines, classification problems and clustering, can be effectively solved using Mahout.

Corporate users of Mahout include Adobe, Facebook, LinkedIn, FourSquare, Twitter and Yahoo.

**Implementing the recommender engine algorithm**

Nowadays, when we shop at online platforms like Amazon, eBay, SnapDeal, FlipKart and many others, we notice that most of these online shopping platforms give us suggestions or recommendations about the products that we like or had purchased earlier. This type of implementation or suggestive modelling is known as a recommender engine or recommendation system. Even on YouTube, we get a number of suggestions related to videos that we viewed earlier. Such online platforms integrate the approaches of recommendation engines, as a result of which the related best fit or most viewed items are presented to the user as recommendations.

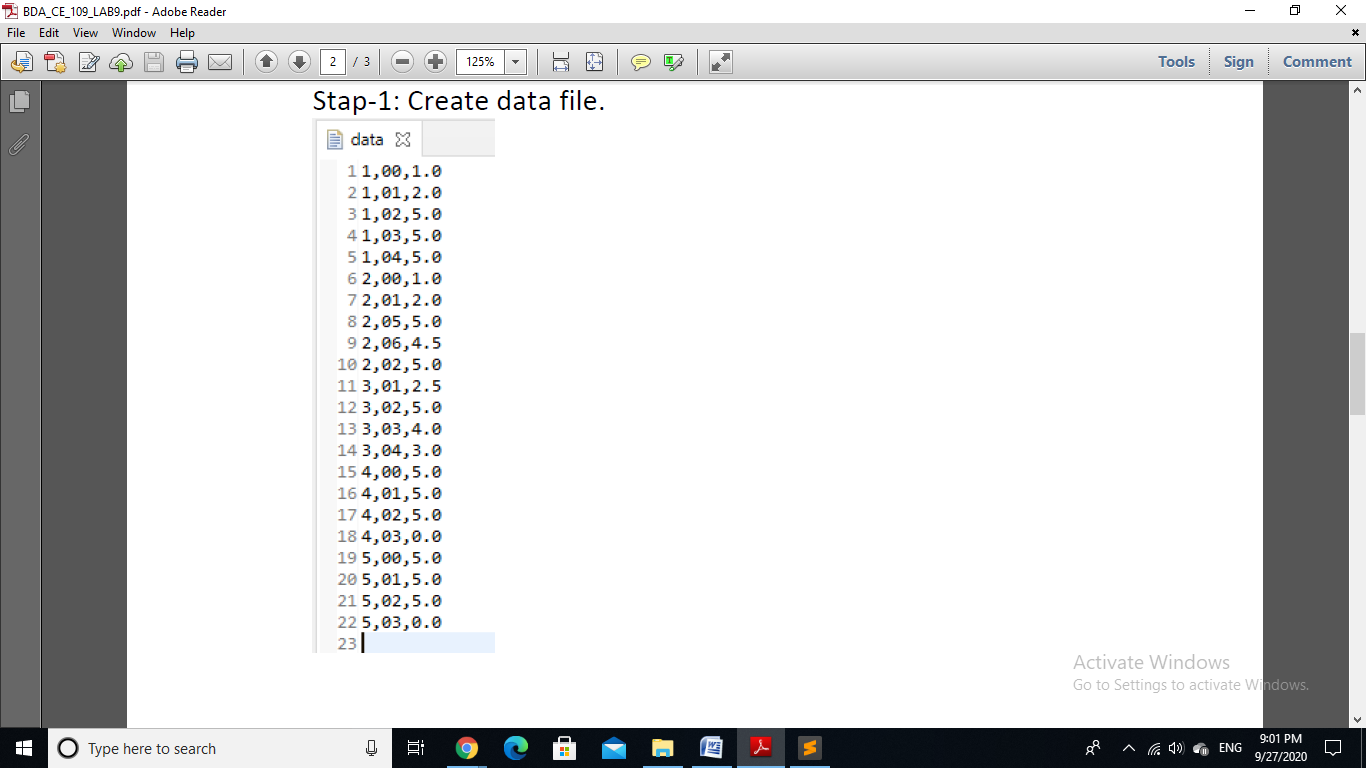
Apache Mahout provides the platform to program and implement recommender systems. For example, the Twitter hashtag popularity can be evaluated and ranked based on the visitor count, popularity or simply the hits by the users. In YouTube, the number of viewers is the key value that determines the actual popularity of that particular video. Such algorithms can be implemented using Apache Mahout, which are covered under high performance real-time machine learning.

For example, a data table that presents the popularity of products after online shopping by consumers is recorded by the companies, so that the overall analysis of the popularity of these products can be done. The user ratings from 0-5 are logged so that the overall preference for the product can be evaluated. This data set can be evaluated using Apache Mahout.

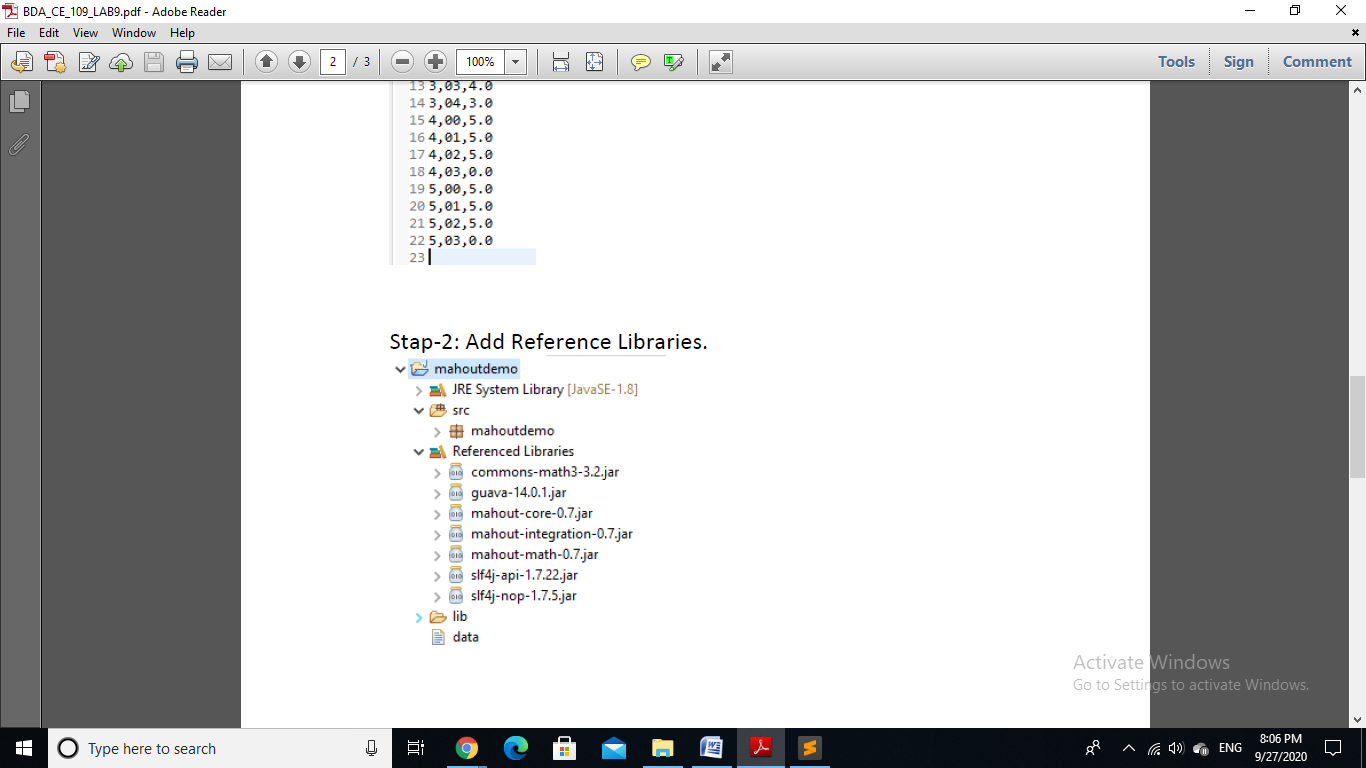
To integrate Java Code with Apache Mahout Libraries on a java IDE, there are specific JAR files that are required to be added.

**Exercise:**

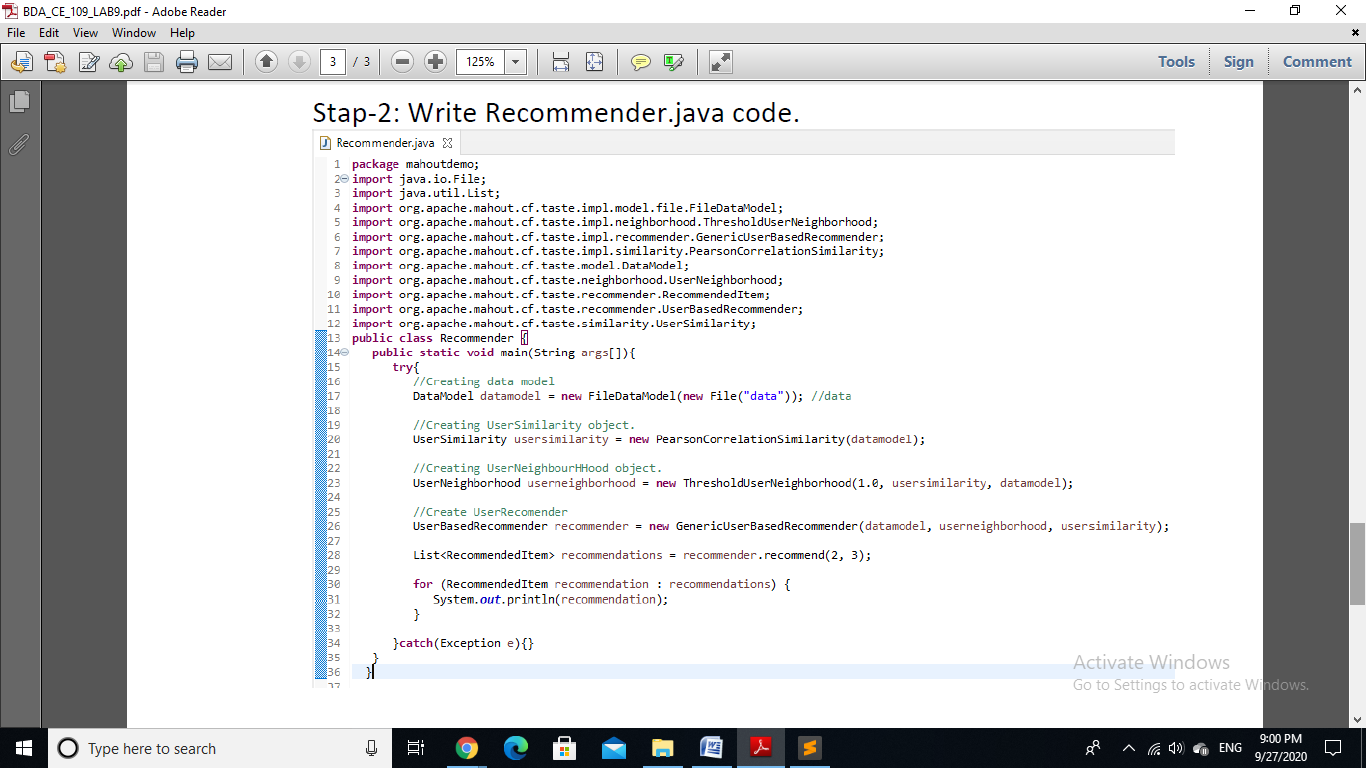
Step-1: Create data file.

****

Step-2: Add Reference Libraries.

****

Step-3: Write below code in Recommender.java file.

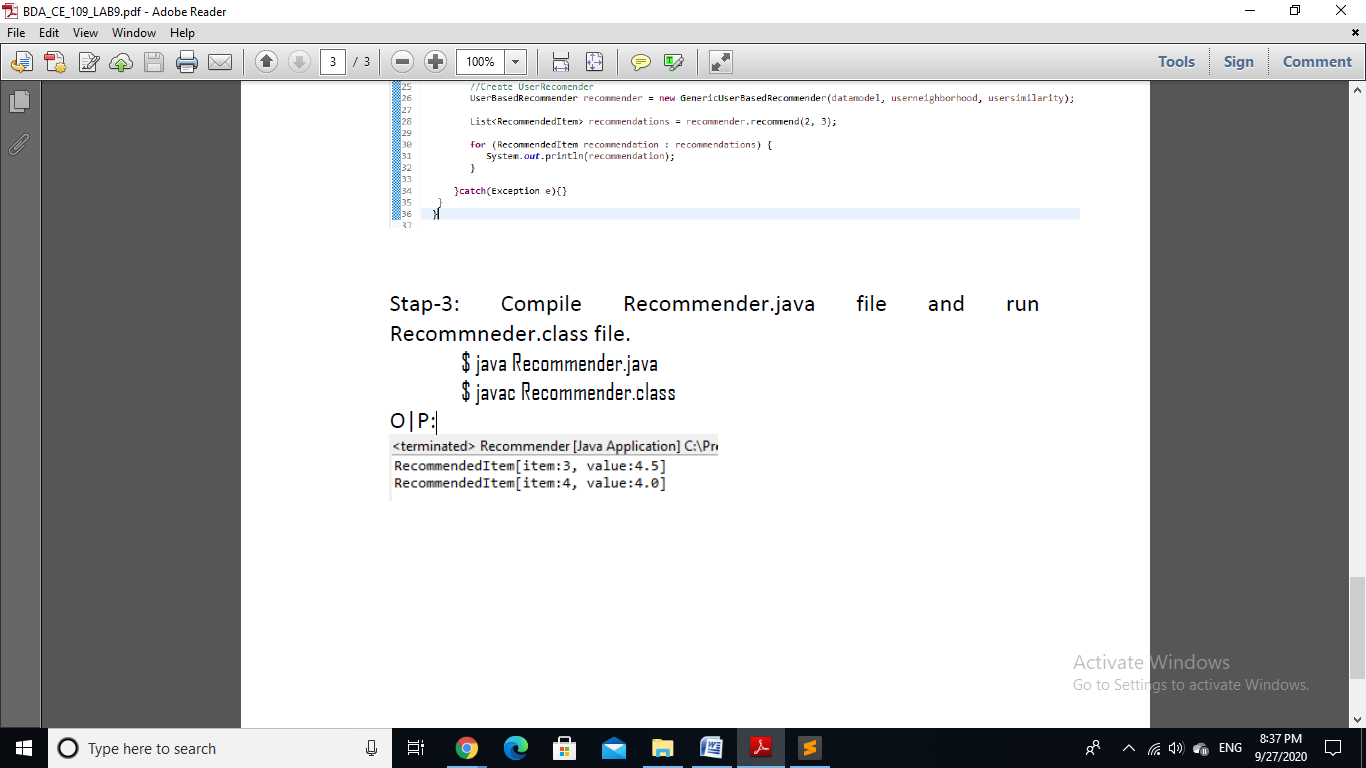
****

Step-4: Compile Recommender.java file and run Recommneder.class file.

$ java Recommender.java

$ javac Recommender.class

O|P:

****