**REPORT FOR ASSIGNMENT 2**

**Question 1 :-**

1. Read the input data from the dataset and map each friend list and user ID to a network RDD.
2. Uploading test data, using a user-defined function to filter out the matched data from the network to TestCases, and storing the outcome in an RDD named User.

A screenshot of a computer

Description automatically generated

1. In the next step we started calculating the strength value based on the network users' IDs' mutual connection.

A screenshot of a computer code

Description automatically generated

1. A screenshot of a computer program

   Description automatically generatedAfter that we started assembling a list of additional userIDs associated with high matches.
2. In the next step we started combining the lists of two connections by first retaining the most often occurring IDs and then adding items from each parent list that are not already in the new combined list.

A screenshot of a computer code

Description automatically generated

1. Define the function responsible for generating suggestions for every test case.
2. Since the result cannot be mapped by applying transformation over transformation, it is saved in a dictionary called dict.

A screenshot of a computer

Description automatically generated

**Question 2:-**

1. A screenshot of a computer

   Description automatically generatedPerusing we enter information from the dataset and map the sentiment of each review into the IMDB\_Dataset RDD.
2. The data was cleaned using a variety of libraries, and the total number of positive and negative reviews was computed. From there, the prior probabilities were counted.

A screenshot of a computer program

Description automatically generated

1. To avoid the divide by zero error, conditional probabilities were computed using Laplace Smoothing, and the result was saved in a dictionary with words serving as the key and conditional probabilities serving as the values.

A screenshot of a computer program

Description automatically generated

1. A screenshot of a computer program

   Description automatically generatedDetermined the likelihoods of a review and used previously determined prior probabilities to classify it.
2. The next step was to text the Naïve Bayes Classifier.

A screenshot of a computer

Description automatically generated

1. At the end after successful testing we got the 75% accuracy using the dataset and the above-described Naïve Bayes classifier.

A screenshot of a computer

Description automatically generated