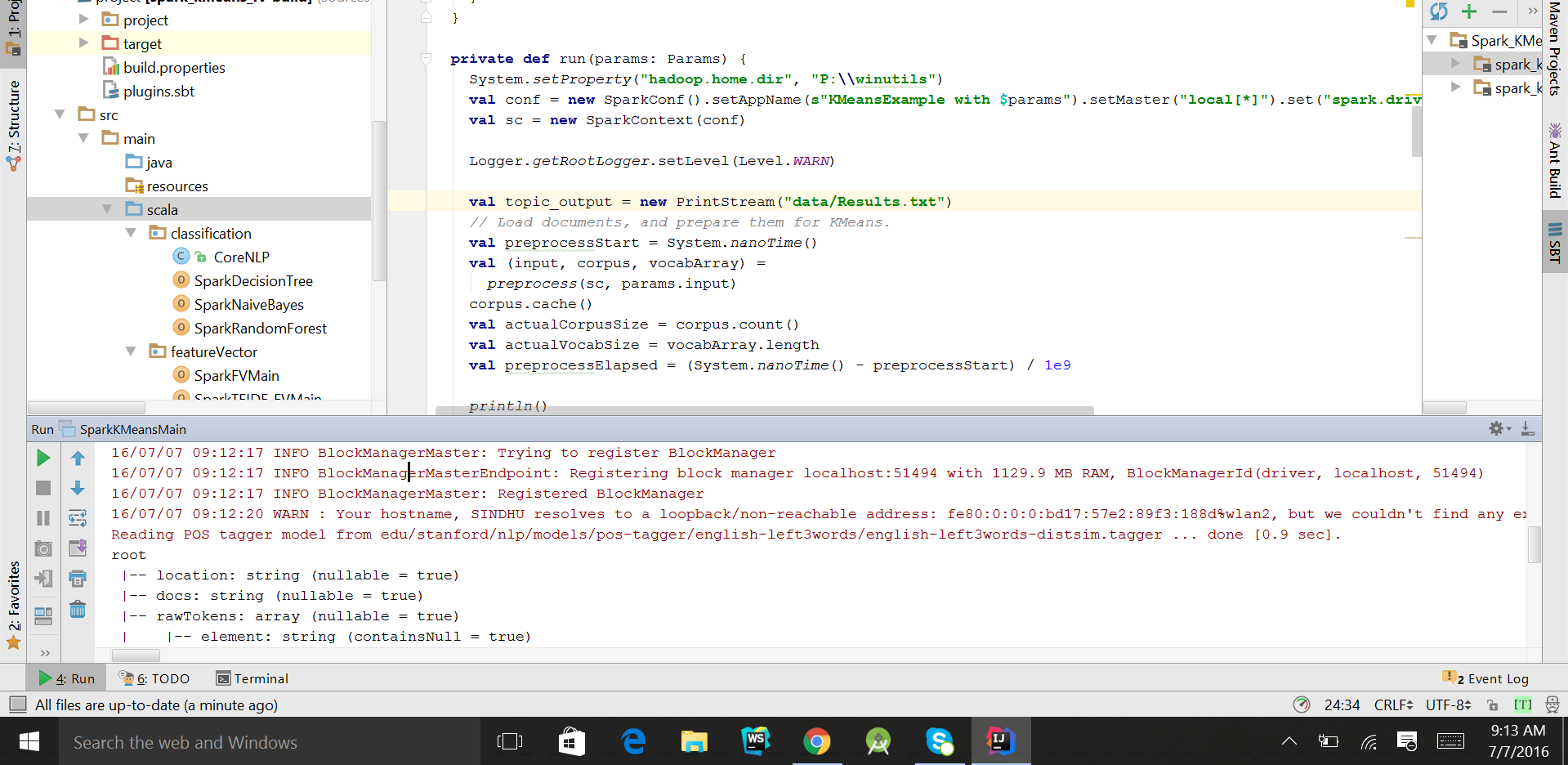
Name: Bhumireddy Ranjitha Reddy

Class Id: 4

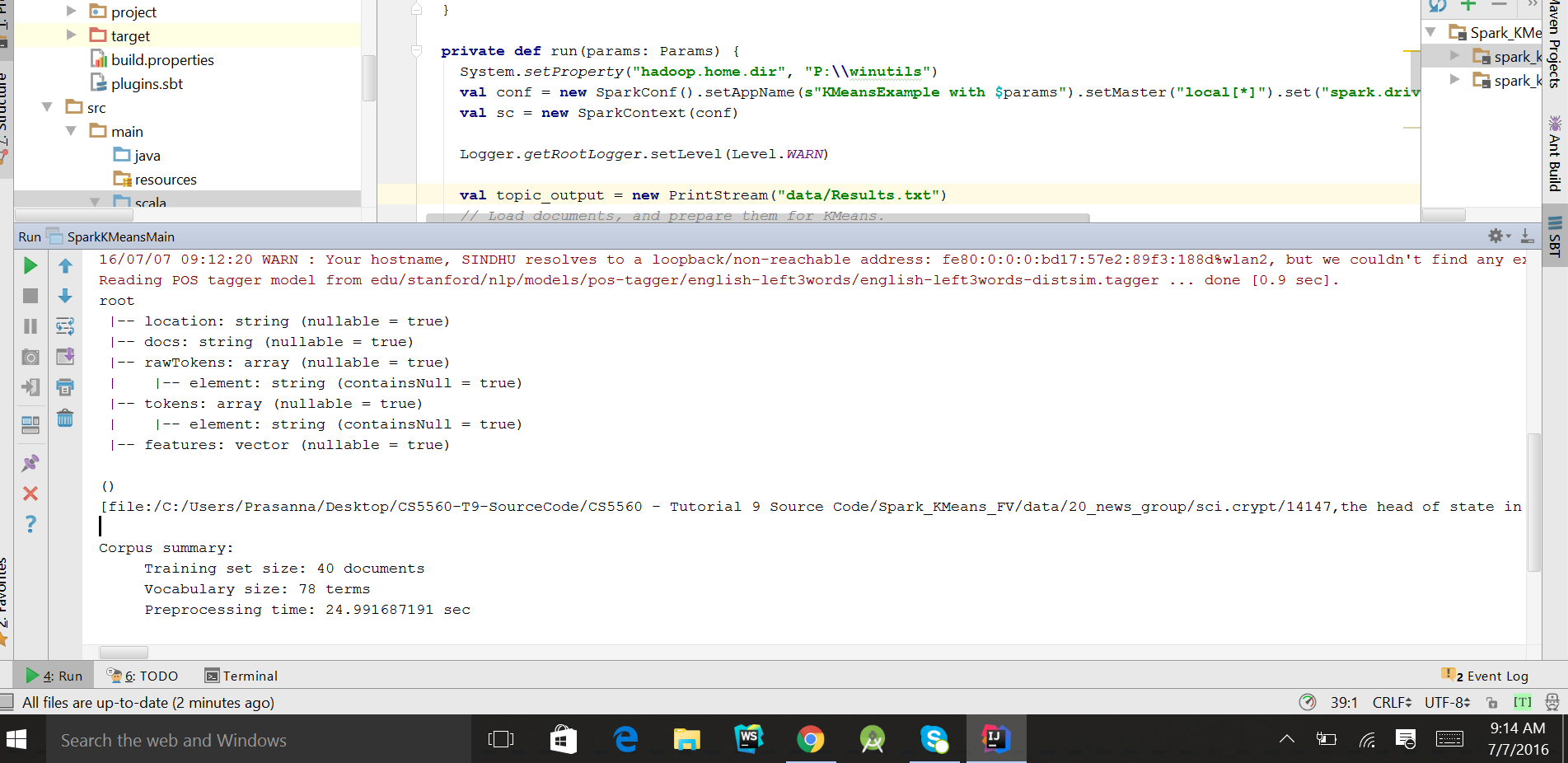
**Input :**

I have given input data from Indian politics.

SPARK KMEANS OUTPUT :



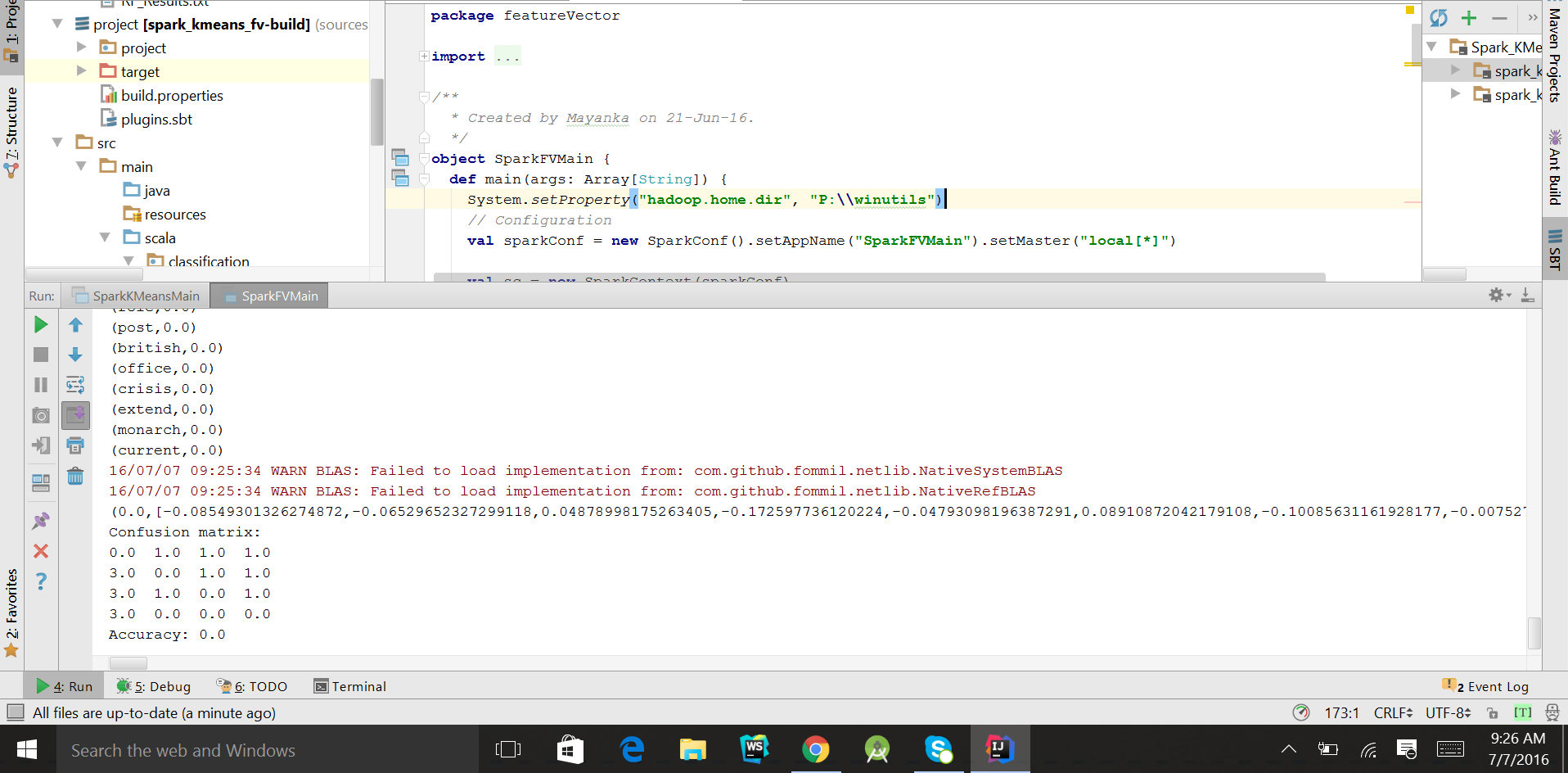
SPARK LDA OUTPUT :



**Difference in the results of SPARK Kmeans and Spark LDA:**

Here clustering of documents can be done by KMeans and Extraction has done by LDA. If both are applied to assign K topics to a set of N documents, the most evident difference is that K-means is going to partition the N documents in K disjoint clusters .On the other hand, LDA assigns a document to a mixture of topics. Therefore each document is characterized by one or more topics.

Feature vector output :



TF-IDF Feature vector output :

