Final Project Team Contribution

Team 15

B-534 Fall 2016

1. **Rohit Nair:**
   - Research:

   Understand the steps to preprocess the data-set [4]

   - Implementation:

   Calculated word frequencies

   Calculated local map of counts of each word

   Aggregate them into one global map

   Calculate word-index map

   Sequential preprocessing calculation completed

2. **Abhijit Karanjkar:**
   - Research:

   Understanding difference in parallelizing mini-batch K_Means

   - Implementation:

   Generating data points

   Random centroid selection

   Mini-batch algorithm skeleton