STEPS TO RUN DOCUMENT SUMMARIZATION CODE IN DATABRICKS

1. Initiate a cluster with the Databricks Runtime Version

**11.3 LTS (includes Apache Spark 3.3.0, Scala 2.12)**

2. Import the notebook into the cluster with the help of the databricks notebook public link

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/78870766149972/4396248947691745/1325961396191934/latest.html>

3. To run summaries for the 5 files we have hosted on github, you need to replace the link of the file in Command 11, Line 4.

The list of the links are given as below

Titanic - https://raw.githubusercontent.com/sai-abhishek123/CS6307/master/Titanic.txt

India - https://raw.githubusercontent.com/sai-abhishek123/CS6307/master/India.txt

Dallas - https://raw.githubusercontent.com/sai-abhishek123/CS6307/master/Assignment\_3/Dallas.txt

Bahubali movie plot - https://raw.githubusercontent.com/sai-abhishek123/CS6307/master/Bahubali%20movie%20plot.txt

Spider Man-Far from home movie plot - https://raw.githubusercontent.com/sai-abhishek123/CS6307/master/Spider%20man%20far%20from%20home%20plot.txt