

# UE20CS581 – Advanced Big Data Analytics

## Assignment-1

**Title: Generating Inverted Index for Mahabharata**

**Professor: Dr. K.V. Subramaniam**

**Submitted by: Suraj S (PES1PG20CS046)**

**Divya M S (PES1PG20CS053)**

**GitHub Repository:**

**Summary:**

- splitted files using csplit command.
- Removing of stop words and cleaning data cleaning.
- Understanding how the MapReduce model works.
- Learning how to set the mapper class and reducer class and updating context for each job.
- Learning about the input and key value pairs.

**Output:**

