

# CS 6350

## ASSIGNMENT 1

Names of student:

Prathamesh Kulkarni – pdk220001

Number of free late days used: 1

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

1. <https://towardsdatascience.com/tf-idf-calculation-using-map-reduce-algorithm-in-pyspark-e89b5758e64c>
2. <https://towardsdatascience.com/tf-term-frequency-idf-inverse-document-frequency-from-scratch-in-python-6c2b61b78558>
3. <https://towardsdatascience.com/understanding-cosine-similarity-and-its-application-fd42f585296a>
4. <https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>

Part-1:

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2458047614741955/2628117978887230/4060260340711020/latest.html>

Part-2:

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2458047614741955/2628117978887247/4060260340711020/latest.html>