CS 6350

Assignment 2

## Name (NetID): Reg Gonzalez (rdg170330)

## Number of free late days used: 0

## Sources/References:

* “Topic 4 - Information Retrieval using MapReduce” Slides & Notes
* “Topic 4 - Apache Spark” Slides & Notes
* “Topic 5 - PySpark Dataframes” Slides & Notes
* “Topic 6 - PySpark MLlib” Slides & Notes
* Dataframes Lab
* MLlib Lab 1
* Dataset for “Part 1 - Friend Recommendation using Mutual Friends” (<https://an-ml.s3.us-west-1.amazonaws.com/soc-LiveJournal1Adj.txt>)
* Dataset for “Part 2 - Implementing Naive Bayes Classifier using Spark MapReduce” (<https://an-utd-python.s3.us-west-1.amazonaws.com/yelp_labelled.txt>)
* Naive Bayes Classifier (<https://www.geeksforgeeks.org/naive-bayes-classifiers/>)
* Naive Bayes from scratch in Spark (<https://ankit-vaghela30.github.io/naivebayes/>)