BU.330.740 Large Scale Computing on the Cloud

**Assignment 2. Sentiment Analysis – pySpark**

Modify Lab 2 Exercise 2 on spam filtering to sentiment analysis.

1. Collect some positive and negative reviews and prepare them as text (.txt) files.
2. Modify the provided PySpark code. You can label positive reviews as **1** and negative reviews as **0**.
3. Implement the code and save all the output in a .ipynb file. Evaluate and analyze your model results using a text cell in the notebook.
4. Name the file **yourname\_assignment2.ipynb** and submit it on Canvas.

**Note:** This assignment is subject to TA interviews.