

MCSD2123 – MASSIVE DATA MINING AND STREAMING

GROUP ASSIGNMENT 2

DURATION: 18 days

DUE: 17th December 2022*

**Presentation on 20th December*

INSTRUCTIONS:

1. Each group is required to identify **ONE (1)** case that relate to **Sentiment Analysis with Machine Learning**
2. Then, identify suitable dataset(s) to be used for the chosen case (you may use more than ONE (1) dataset if needed).
3. Study and perform EDA on the datasets.
4. Explore and select **THREE (3)** supervised machine learning algorithms for training the sentiment models using the identified dataset(s). Make sure datasets are properly prepared and processed.
5. Evaluate and compare the performance of all models.
6. Conclude your findings

DELIVERABLES:

1. Reporting Slides (in PDF)
Your slides should cover following topics**:
 - i. Title
 - ii. Executive Summary
 - iii. Introduction
 - i. Background of the case
 - ii. Problems you want to find answers
 - iv. Dataset
 - v. Methodology
 - i. Overview
 - ii. Preprocessing
 - iii. Model Training
 - iv. Model Evaluation
 - vi. Results and Findings
 - vii. Conclusions

*** You may use more than 1 slide for explanation of any of the mentioned topics above.*

2. Python scripts/ notebooks / saved project files from analytics tools
3. Document that states the distribution of work among group members. (in PDF)