

**SCHOOL OF COMPUTER TECHNOLOGY**

AASD 4001 Mathematical Concepts for Machine Learning

**Term Project:**

1. Obtain a large classification database. The database needs to have at least 2 classes, more than 5 features and over 200 samples. Each group has to use a unique dataset.
2. Provide a description of the dataset used including explanation of various features.
3. Pre-process and clean the dataset as appropriate.
4. Use following approaches for classification of the dataset:
   1. Logistic Regression
   2. Decision Tree
   3. Random Forest
   4. SGD
   5. SVM
5. Use GridSearchCV to tune the parameter of each of the above models. Can you obtain better results in this step for any of the models? Discuss your observations.
6. Randomly (or based on certain hypothesis) remove some features and re-evaluate the models. Document your observations with respect to models performances.
7. Present your work including approach and findings during the class on January 31, 2023. Each group will have a maximum of 15 minutes to present their project. It is advised that your PowerPoint files to be no longer than 15 slides.
8. Prepare a written technical report of no longer than 15 pages to discuss the problem statement, various steps conducted, summary of findings and conclusions. Submit the report and the notebook file (with proper headings, explanatory comments and code sections) by the midnight of February 4th, 2023.