

**SCHOOL OF COMPUTER TECHNOLOGY**

AASD 4001 Mathematical Concepts for Machine Learning

**Project**

**Random Forest Model**

In this exercise students are expected to build either a

1. Random Forest Classifier
2. Random Forest Regressor

Note that just one of these items are required.

**Database:**

Students are free on the choice of the specific task they like to do or the databases they like to analyze.

The database need to meet the following criteria:

1. Database need to have at least 2 classes, but more classes are recommended, like 5 or more.
2. Database should at least have 300 samples/rows (1000 or more is recommended)
3. Database should at least have 8 columns/features (15 or more is recommended)

**Report:**

Put your results in a report. Your report needs to include the following sections:

* The Problem statement,
* The Database,
* The model you picked to solve the problem,
* Results, the model performance (test, valid), the loss, predictions…

(like use of confusion matrices etc…)

* Conclusions

In your results please comment and discuss the followings:

1. Evaluate the model, how?
2. How your model change when the number of estimators (decision trees) changes?
3. What is the best number of estimators? How you can select the best number of estimators?

Important Note:

1. Make sure to document your findings and back up your conclusions using the results, use visualization.
2. Note that you have to present your progress/results on Friday Oct 1st, but the final delivery date is Oct 5th.

If Oct 1st is not work for you, you can present on Sep 28th or 30th, Please notify me in advance.