3684 – Advanced Topics in Machine Learning, Spring 2022
Home Assignment #1a – Ensemble Learning practical work

Lecturer: Dr. Leon Anavy
Teaching Assistant: Mr. Alon Oring

General instructions:

1. Submission is **individual.**
2. Submission must include python code **and** a written report.
3. You may use external libraries. Specify all required libraries in a proper manner.
4. Your code must be reproducible. Code that will not run will result in a grade reduction.
5. Your report should be clear, coherent, and concise. The report should not exceed 10 pages.
6. Invest thoughts and considerations to the way you choose to present data and experimental results.
7. All figure and plots should include captions, labels and data units. Pay attention to data visualization guidelines.
8. Make sure to use correct ML methodologies and justify your selections (split the data to train/test, tune hyperparameters, report relevant performance measures).

Assignment tasks:

The goal of this assignment is to practice the use of ensemble methods for classification problems.

1. Choose one of the following datasets:
   - Telco customer churn
   - Default of credit card
   - Hotel booking demand
2. Explore the dataset and its properties using standard Exploratory Data Analysis methods
3. Define a classification problem based on the chosen dataset.
4. Perform analysis of the performance of different classification algorithms on the dataset. (Decision trees, Random Forests, Bagging, Boosting, etc.).
5. Demonstrate different ML concepts that were covered in class using the dataset (Overfitting, Generalization, Bias/Variance, etc.)

Summarize all your work in a scientific/professional report.

Class presentation:

If you have chosen this assignment as your class presentation assignment you are required to prepare a 30 minutes presentation in which you will need to showcase your work. You should cover all aspects of your work in the presentation.