Assignment 5: Data Lab, Random Forest

- Due by 27th October, 2022 before 5pm IST.
- To be submitted to the following email address: office.of.gr@gmail.com
- The subject of the email should be: Assignment Number [5]: Data Lab, 2023
- Please clearly mention your name and roll number.
- Submit your work as a single pdf file. Additional material, code, etc can/should also be submitted, but there should be atleast 1 pdf, which has the entire assignment.
- Wherever there is code, in the assignments, the code should be well documented and easy to understand / follow.

The objective of the assignments is three fold. One is to be able to develop expertise in writing and communicating about technical topics. This will be done by using the IEEE conference style format for all assignments. The other is to explain, in your own way, the mathematical ideas that are embedded within the technical topic of interest. For example, in this case it is random forest. The third is to use the topic, in this case of random forest, to understand a problem from the real world. So in a sense the objective is to write what one may call a mathematical essay on random forest Classifier.

Title could be: Assignment 5: a mathematical essay on random forest.
Abstract: Give a brief overview of your assignment.
Author: Name, Department, Institution, Email

Section 1: Introduction

In this section, the 1st paragraph should be on a broad overview of the topic. The 2nd paragraph should be an overview of the technical aspects (i.e. in this case it is a random forest). The 3rd paragraph should be about the problem that you are aiming to solve/understand using random forest. Finally, the 4th paragraph should give an overview of the paper.

Section 2: Random Forest

This section should outline the key principles underlying random forest.

Section 3: Data

The *Car Evaluation Database* was derived from a simple hierarchical decision model. The prediction task is to classify a car based on its safety. The data is provided in the car_evaluation.csv file.

Variable	**Definition**	**Key**
buying	buying price	vhigh, high, med, low
maint	Price of the maintenance	vhigh, high, med, low
doors	Number of doors	2, 3, 4, 5, more
persons	Capacity in terms of persons to carry	2, 4, more
lug_boot	The size of luggage boot	small, med, big

safety	Estimated safety of the car	low, med, high
Target	Target variable to predict	unacc, acc, good, vgood

Section 4: The problem

(a) Outline the problem, and plot/visualise the data.
(b) Make progress on the problem, by applying the techniques of random forest to the problem at hand.
(c) Discuss any insights and observations.

Section 5: Conclusions

Write about 1 paragraph on the key insights that were obtained from your study and also outline any further avenues of possible investigation.

References

Please put in all the references that you have used during the assignment. The format should be the same as the IEEE conference format.