

## Problem 2: Air Purification and Plant Health in Smart Greenhouses

We aim to investigate how air purification systems and automated watering influence two physiological indicators of plant health in smart greenhouses. The dataset `greenhouse.txt` contains measurements of:

- `chlorophyll`, the average chlorophyll concentration in leaves, measured in  $\mu\text{g}/\text{cm}^2$ ;
- `growth`, the average daily height increase of plants, measured in  $\text{cm}/\text{day}$ .

The data come from 200 greenhouse units, each growing the same species of lettuce under controlled conditions. Each unit is tagged with two binary variables:

- `air_purifier`, indicating whether the unit was equipped with an air purification system (1 = yes, 0 = no);
- `auto_watering`, indicating whether automated watering was used (1 = yes, 0 = no).

- a) Do air purification and automated watering systems have a significant effect on the two plant health indicators? Support your answer using a MANOVA model.
- b) Clearly state the assumptions of the MANOVA model used in (a), and assess their validity using appropriate diagnostics.
- c) Based on your analysis in (a) and (b), would you consider modifying the model? Explain.
- d) Build an appropriate discriminant classifier to predict whether the lettuces of a specific greenhouse were grown using an air purification system, based on the chlorophyll concentration and the average daily growth. Similarly, develop a classifier to predict whether automated watering was used. Report the estimate of the actual error rate with leave-one-out cross validation for the two classifiers built.
- e) Given the MANOVA results, would you consider a different approach for predicting whether both air purification and automated watering were used?

Upload your results here: <https://forms.cloud.microsoft/e/6Z4GLuB3Fz>