



# AIML

# MODULE PROJECT

Computer Vision

TOTAL  
SCORE

30

General Instructions:

- 1. Submission of all the parts is expected in 1 notebook only
- 2. Expected submission format: 1 '.ipynb' notebook and 1 '.html' notebook only
- 3. 50% marks will be deducted if insights/steps are missing in the corresponding questions.
- 4. If output for any code cell is missing, 50% marks will be deducted.
- 5. Any kind of Plagiarism will lead to 0 (zero) Marks.

Submission Format:

- 1. '.ipynb' (Jupyter Notebook) and
  - 2. '.html' (Jupyter Notebook > File > Download as > HTML)
- 5 Marks will be deducted if submission in any of the formats is missing.

- **DOMAIN:** Botanical Research
- **CONTEXT:** University X is currently undergoing some research involving understanding the characteristics of plant and plant seedlings at various stages of growth. They already have have invested on curating sample images. They require an automation which can create a classifier capable of determining a plant's species from a photo.
- **DATA DESCRIPTION:** The dataset comprises of images from 12 plant species.  
Source: <https://www.kaggle.com/c/plant-seedlings-classification/data>.
- **PROJECT OBJECTIVE:** To create a classifier capable of determining a plant's species from a photo.

Steps and tasks: [ Total Score: 30 Marks]

- 1. Import and Understand the data [12 Marks]
  - A. Extract 'plant-seedlings-classification.zip' into new folder (unzipped) using python. [2 Marks]  
*Hint: You can extract it Manually by losing 2 marks.*
  - B. Map the images from train folder with train labels to form a DataFrame. [6 Marks]  
*Hint: Create a DataFrame with 3 columns: Name of image, Species/class/type of image & actual image..*
  - C. Write a function that will select n random images and display images along with its species. [4 Marks]  
*Hint: If input for function is 5, it should print 5 random images along with its labels.*
- 2. Data preprocessing [8 Marks]
  - A. Create X & Y from the DataFrame. [2 Marks]
  - B. Encode labels of the images. [2 Marks]
  - C. Unify shape of all the images. [2 Marks]
  - D. Normalise all the images. [2 Marks]
- 3. Model training [10 Marks]  
*Checkpoint: Please make sure if shape of X is (No.of images, height, width, No. Of channels). If not, you need to correct it otherwise it will be issue during model training.*
  - A. Split the data into train and test data. [2 Marks]
  - B. Create new CNN architecture to train the model. [4 Marks]
  - C. Train the model on train data and validate on test data. [2 Marks]
  - D. Select a random image and print actual label and predicted label for the same. [2 Marks]