



DEPARTMENT OF COMPUTER SCIENCE

COMP338 - Artificial Intelligence

Course Project 2

Due date: Jun 13, 2025

Dr. Mohammad Helal

Dr. Radi Jarrar

Dr. Abdel-Rahman Hamarsheh

Project Title:

Classification by Decision Tree to Identify Most Suitable Crops for Certain Environment

Objective:

Using the programming language of your choice and using the dataset provided below to train a Decision Tree in order to identify the best suitable crop for a given environment. The dataset serves a reasonable training, testing and evaluation set as it contains 7 parameters representing the climatic conditions for a large number of farms, and an output label for a particular crop that suits each environment.

Use the **5-folds cross-validation** procedure to validate your decision trees, in other words, you need to redo the training process 5 times and then compare: **Accuracy, Precision and Recall** between each iteration.

After training is complete, you need to build a platform that helps users enter the climatic parameters for the real environments then apply them to the Decision Tree in order to recommend the most suitable crop for that particular environment. Your program should also present the final Decision Tree in a graph.

The Dataset:

The dataset provides real life data for a number of different farmlands, each labeled with the most suitable crop to grow that suits that environment.

The Features included in the dataset:

Features (Inputs):

- N – Nitrogen content in the soil (in mg/kg)
- P – Phosphorus content in the soil (in mg/kg)
- K – Potassium content in the soil (in mg/kg)
- temperature – Average temperature in °C
- humidity – Average relative humidity in %
- ph – Soil pH value
- rainfall – Rainfall in mm

Label (Output):

- label – The crop most suited to the given conditions (e.g., rice, maize, cotton, etc.)

Link to the dataset:

<https://www.kaggle.com/datasets/madhuraatmarambhagat/crop-recommendation-dataset>

What to Submit:

Each team is expected to submit a working implementation of the algorithm and a Report.

Implementation and Experimental work should include:

- Perform the 5-fold cross-validation
- Calculate Accuracy, Precision, and Recall for each fold (or iteration)
- Generate and display the decision tree(s)
- Allow user to enter new parameters then use the tree to present a recommended crop suitable for these parameters

Your report should include:

- Decision Tree algorithm
- overview on the dataset
- explain Accuracy
- Precision and Recall
- present your results

- discuss the difference in performance measurements between the folds
- explain your source code
- present the decision tree.

Grading Criteria:

- Grades are distributed as follows:
Implementation: 33%
Report: 33%
Discussion: 33%
- For the project:
 - proper implementation of the algorithm
 - calculation of Accuracy, Precision and Recall
 - implementation of 5-fold cross-validation
 - generating the decision tree(s)
 - implementing the tree
- For the report:
 - explanation of necessary algorithms and evaluation metrics
 - overview on the dataset
 - explanation of the code
 - discussion of the experimental work, specially the difference in evaluation metrics between the 5-folds
 - final results and conclusion
- Discussion:
You will be asked several questions some to verify your contribution to the project implementation and some to grade your understanding of the material itself
Each student will be graded separately, so team members may have different scores.

Tools and Technologies:

- Programming Language: Java, Python, R, Octave
- Libraries: Weka or NumPy, Matplotlib (for plotting)

Important Rules:

- Late submissions will not be graded reply before midnight of 13th of June
- The project is a teamwork, each team consists of two students
- Team members can be from different sections
- There will be a project discussion and it will have 33% of the project scores
- Plagiarism is not tolerated and can lead to failing the whole course