COS40007 Artificial Intelligence for Engineering

Portfolio Assessment 1: "Hello Machine Learning for Engineering"

Due: by Sunday of Week 3 (23/03/2025 23:59 PM) in Canvas

AIM

This task aims to demonstrate your familiarity with Python programming, your favourite IDE, engineering datasets and programming machine learning with Python. The task is for you to understand data exploration, data labelling, feature engineering and machine learning model development.

Tasks

You must complete a portfolio assessment based on Studio 1 and 2 activities. Here are the tasks you must complete for your selected dataset in Studio 1.

1) Define 3 to 5 class labels in the target variable of your dataset if this contains a numerical value. Otherwise, if the target variable is categorical, leave it as it is. Make sure there is a nearly balanced class distribution when defining class labels.
2) Conduct normalisation of numerical features and integer value categorisation of other types of features.
3) Apply feature engineering and create new features (with possible composite features) based on your outcome in EDA.
4) Create and test decision tree models with at least five sets of features you derived from the EDA summary and compare them. Then, put the comparison results in a table.

Checklist

The portfolio assessment submission should be a document (word or pdf) with the following.

- Your name and Student number
- The studio class you attend (for example, you attend Studio 1-1, then write Studio 1-1)
- The dataset you selected in Studio 1 (**0.5 mark**)
- Briefly Explain the reason for your choice (**0.5 mark**)
  [for example, I am an electrical engineering major, so I am interested in exploring the combined cycle power plant dataset]
- A summary of your exploratory data analysis (EDA) in Studio 1 **(1 mark)**
- Class labelling for target variable / developing ground truth data **(0.5 mark)**
- Feature engineering and Feature selection **(2 marks)**

- Training and decision tree model development **(2 marks)**
- Final comparison table **(2 marks)**
- A brief summary of your observation in the comparison table **(0.5 mark)**
- Appendix: place your source code/notebook in One Drive or Google Drive and provide us with a shared link of your source code **(1 mark)**

Total **10 marks**