COS40007 Artificial Intelligence for Engineering

Portfolio Assessment-1: "Hello Machine Learning for Engineering"

Due: by Sunday of Week 3 (11/08/2024 23:59 PM) in Canvas

AIM

The aim of this task is for you to demonstrate your familiarity with python programming, your favorite IDE, engineering dataset 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 need to complete a portfolio assessment based on activities in Studio 1 and 2. Here are the tasks you need to 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 numerical value. Otherwise, if the target variable is categorial then leave this as it is. Make sure there is nearly balanced distribution of classes when you are defining class labels .
2) Conduct normalisation of numerical features and integer value categorisation of other type of features
3) Apply feature engineering and create new features (with possible composite features) based on your outcome in EDA
4) Crate and test decision tree models with at least 5 different set of features you came up from the summary of EDA and compare them. Put the comparison result in a table.

Checklist

The portfolios 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 of your choice (**0.5 mark**)
  [for example, I am in electrical engineering major so I am interested to explore 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 onedrive or google drive and provide us a shared link of your source code **(1 mark)**

Total **10 marks**