PRACTICAL ASSIGNMENT

Module
[D4A Essential]

Module Code
AI43005FP

Duration
8 hours

Title: Perform AI data collection, AI training and inference

Tools, Equipment and Materials:

- Computer with Internet access
- Documentation & Presentation Software

Instructions:

1. You are to perform AI data collection, AI training and inference of a case study individually.

2. Analyse and explore the case study provided.

3. There are six parts to this assignment. Problem definition, data, evaluation, features, modelling and experiments.

4. You are to map it to AI project cycle – Problem scoping, Data acquisition, Data exploration, Modelling, Evaluation and Deployment.

Deliverables:

You have to submit one complete set of items and present the assignment to demonstrate the application to the Lecturer upon completion of the assignment.

You have to zip and submit one set of items consists of

a. Python code,
b. Presentation slides.
c. There should be only one submission per person.
d. Ensure that the following contents are included in the presentation slide.

1	Cover slide including project title and student name.
2	Problem statement using 4Ws problem canvas.
3	Problem definition, data, evaluation, features, modelling and experiments.

PRACTICAL ASSIGNMENT

Module
[D4A Essential]

Module Code
AI43005FP

Duration
8 hours

Data for AI Essential Scenario

Predicting Heart Disease using Machine Learning

In our case, the problem we will be exploring is binary classification (a sample can only be one of two things).

This is because we're going to be using a number of different features (pieces of information) about a person to predict whether they have heart disease or not.

In a statement,

Given clinical parameters about a patient, can we predict whether or not they have heart disease?

As such, we'll look at the following topics.

- Exploratory data analysis (EDA) - the process of going through a dataset and finding out more about it.
- Model training - create model(s) to learn to predict a target variable based on other variables.
- Model evaluation - evaluating a models predictions using problem-specific evaluation metrics.
- Model comparison - comparing several different models to find the best one.
- Model fine-tuning - once we've found a good model, how can we improve it?
- Feature importance - since we're predicting the presence of heart disease, are there some things which are more important for prediction?
- Cross-validation - if we do build a good model, can we be sure it will work on unseen data?
- Reporting what we've found - if we had to present our work, what would we show someone?
- To work through these topics, we'll use pandas, Matplotlib and NumPy for data analysis, as well as, Scikit-Learn for machine learning and modelling tasks.

Instruction to students:

Please answer ALL the questions. You may refer to online documentation (e.g. keras, pandas, etc) for help on syntax. No copying of codes wholesale is allowed.

The assignment will require the following steps:

PRACTICAL ASSIGNMENT

Module
[D4A Essential]

Module Code
AI43005FP

Duration
8 hours

1. Problem scoping.
   1.1. Defining problem statement using 4Ws (Who, What, Where, Why) problem canvas from the project scenario.
2. Data acquisition and Data exploration.
   2.1. Preparing the tools by importing libraries.
   2.2. Loading the data.
   2.3. Data exploration (exploratory data analysis)
      2.3.1. Comparing two columns.
      2.3.2. Correlation between independent variables.
3. Modelling.
   3.1. Train, test and split.
   3.2. Model choices: logistic regression, K-nearest neighbors and random forest.
   3.3. Model comparison.
   3.4. Hyperparameter tuning and cross-validation.
4. Evaluation and Deployment.
   4.1. ROC curve and AUC scores.
   4.2. Confusion matrix.
   4.3. Classification report.
   4.4. Feature importance.
   4.5. Experimentation.

- END-