# HIT 391 – ASSIGNMENT 2 (Individual Coding Exercise) Sem 1, 2025

Welcome to the second assignment of this course. It requires to submit the Machine Learning coding exercise, consisting of a 500-word report and two coding files. This assignment consists of 4 datasets related to healthcare. From these datasets you should select any 2 and perform both Task 1 and Task 2.

Task 1: Classification:

- Step 1: Loading Data, Data Pre-processing, EDA
- Step 2: Feature Engineering, Creating Train, and Test Datasets
- Step 3: Apply at least 2 algorithms for classification (Training and Testing)
- Step 4: Generate at least 2 Evaluation Metrics on each algorithm.
- Step 5: Comparing the results.
- Step 6: Fine Tune the best algorithm.

Task2: Regression

- Step 1: Loading Data, Data Pre-processing, EDA
- Step 2: Feature Engineering, Creating Train, and Test Datasets
- Step 3: Apply at least 2 algorithms for regression (Training and Testing)
- Step 4: Generate at least 2 Evaluation Metrics on each algorithm.
- Step 5: Comparing the results.
- Step 6: Fine Tune the best algorithm.

Datasets:

D1: Predicting diseases from symptoms.

D2. Predict whether a patient is likely to get a heart stroke.

D3. Predicting medical insurance costs

D4. Predicting life expectancy using WHO data.

You have 2 scenario options (try to choose the best datasets for these scenarios):

- You can implement Task 1 and Task 2 on a single dataset.

**Or**

- You can implement Task 1 on one dataset and Task 2 on other.

## Assignment Submission

**You should submit 3 documents.**

- A single file with all the steps for Task 1(.py or. ipynb). (5 marks)
- A single file with all the steps for Task 2(.py or .ipynb). (5 marks)
- A 500 words report explaining about the data, different steps, and their results by using graphical charts. The report should be concise and provide some meaningful insights. (10 marks)

**Submission Format**

The files should be named as illustrated below.

- If you choose D1 or D2 or … for Task 1, then the name should be '11.py' or '21.py' or … Similarly, if you choose D1 or D2 or …. for Task 2, then the name should be '12.py' or '22.py'.
- The report (.doc or pdf) can have any name.