Coding Assessment for the SingCLOUD Project

Problem Statement:

The MIMIC-III (Medical Information Mart for Intensive Care) Clinical database consists of anonymized health-related data of more than 40,000 patients at the critical care units of Beth Israel Deaconess Medical Center from 2001 to 2012[1].

In this assessment, we are interested in data exploration of the MIMIC-III dataset and prediction of hospital mortality. We will be using the dataset hosted on Kaggle[2].

Task:

Your tasks are as follows:

1. Perform exploratory data analysis (EDA) on the given dataset
2. Develop a machine learning model to predict mortality of ICU patients (you may develop more than one model for comparison).
3. Address the following questions:
    a. Which are the most significant predictors of mortality among ICU patients?
    b. Are there correlations between variables that significantly affect mortality prediction?
    c. Explain in detail how you prepared the data for this machine learning task, and the rationale behind the steps taken.
    d. Which model did you choose to predict mortality of ICU patients and why do you think it is suitable? How well did this model perform?
4. Prepare a 10 min presentation to address the questions as well as any other findings. Be prepared to show your code and explain any questions on it during Q&A.

Evaluation Criteria:

We will be assessing you on the following criteria:

Coding:

- Proposed method is an acceptable solution for the problem statement.
- Demonstrates understanding in handling data.
- Code is neat, readable, and well documented.

Presentation:

- Clear and concise, able to explain clearly how the proposed method addresses the questions.
- Overall structure, layout, and flow of the presentation.

---

[1] Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035.

[2] https://www.kaggle.com/datasets/saurabhshahane/in-hospital-mortality-prediction

- Adhere to the time limit of 10 mins.
- Able to handle questions well during Q&A.

Submission Guidelines:

You may use any suitable Python libraries of your choice. Do note that all code has to be written by you. The deadline for submitting the code is **1 week after release of the assessment.** Please upload your code and documentation (if any) to Github and email us the link. You will present your work **2 weeks after release of assessment** to the group.