

# TRUNG TÂM TIN HỌC ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP. HỒ CHÍ MINH

## Đề thi giữa kỳ:

### MACHINE LEARNING WITH PYTHON

**Thời hạn nộp bài: 23h30 ngày 19/01/2022 (Thứ 4)**

\*\*\* HV tạo 1 thư mục **LDS6\_k272\_HoVaTen\_Giua\_ky** trong thư mục  
**Đã chia sẻ** trên Google Drive đã share, lưu tất cả bài làm vào để chấm điểm \*\*\*

\*\*\* HV sẽ bị trừ điểm nếu bài làm giống nhau \*\*\*

\*\*\* HV phải gửi mail đính kèm link của thư mục

**LDS6\_k272\_HoVaTen\_Giua\_ky** đúng hạn nộp bài, sau hạn nộp bài nếu HV không gửi thì sẽ không được chấm điểm \*\*\*

### Chú ý:

- Mỗi câu là một file (có thể viết trên jupyter notebook/ jupyter lab/colab) □ HV cần thực hiện việc tiền xử lý dữ liệu cho từng câu.
- Mỗi câu đều phải đưa ra nhận xét, giải pháp cho các lựa chọn. Nhận xét viết dưới định dạng Markdown.
- Các phần trực quan hóa thì vừa phải trực quan vừa phải giải thích.

#### 1. **Classification**

- Cung cấp bộ dữ liệu Ecoli, tập tin **ecoli.csv** trong folder đề thi (thông tin chi tiết và dữ liệu cũng có thể download trên <https://www.openml.org/d/1011>) với 8 features và bảng kết quả run (<https://www.openml.org/t/3874>) đã được thực hiện để tham khảo.
- Yêu cầu :
  - Hãy chọn thuật toán phù hợp để xây dựng model dự đoán một mẫu có **binaryClass** là positive("P") hay negative ("N"). Giải thích lý do tại sao bạn chọn thuật toán này?
  - Chú ý: Thực hiện tất cả các bước để giải quyết vấn đề này từ tiền xử lý dữ liệu => báo cáo kết quả.

#### 2. **Regression**

- Cung cấp bộ dữ liệu BMW, tập tin **bmw.csv** trong folder đề thi (thông tin chi tiết và dữ liệu cũng có thể tham khảo và download trên <https://www.kaggle.com/adityadesai13/used-car-datasetford-and-mercedes?select=bmw.csv>).
- Yêu cầu :
  - Hãy chọn thuật toán phù hợp để xây dựng model dự đoán giá xe (**price**). Giải thích lý do tại sao bạn chọn thuật toán này?
  - Chú ý: Thực hiện tất cả các bước để giải quyết vấn đề này từ tiền xử lý dữ liệu => báo cáo kết quả.

**--- Chúc các bạn làm bài tốt ---**