



fit@hcmus

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG - HCM
KHOA CÔNG NGHỆ THÔNG TIN

MÔN HỌC: MÁY HỌC

ĐỒ ÁN THỰC HÀNH I

1 Nội dung đồ án

1.1 Giới thiệu bài toán

Chi phí sử dụng dịch vụ y tế là bài toán nhằm xác định các yếu tố ảnh hưởng đến số tiền mỗi cá nhân phải chi trả cho hoạt động y tế cá nhân để từ đó dự đoán chi phí sử dụng y tế của 1 người dựa trên các yếu tố đã thu thập. Cụ thể, sinh viên được cung cấp 2 tập tin “train.csv” và “test.csv” chứa thông tin y tế cá nhân tại Mỹ được thu thập dựa trên các thông tin được thu thập sau:

- **Age:** Tuổi
- **Sex:** Giới tính
- **BMI:** Chỉ số khối cơ thể
- **Children:** Số lượng trẻ con/người phụ thuộc
- **Smoker:** tình trạng hút thuốc
- **Region:** khu vực sinh sống
- **Charges:** Chi phí y tế cá nhân

1.2 Mục tiêu

Tìm hiểu và vận dụng các kỹ thuật:

- **Nhận diện bài toán:** Sinh viên cần tìm hiểu bài toán và dữ liệu được giao nhằm xác định nội dung và ý nghĩa bài toán thực tế cần giải quyết. Thông qua đó, sinh viên có khả năng ánh xạ vấn đề thực tế sang bài toán lập trình.
- **Giải quyết vấn đề:** Sinh viên được yêu cầu đưa ra các giải pháp và hướng tiếp cận nhằm giải quyết được yêu cầu bài toán thực tế.
- **Xử lý và phân tích dữ liệu:** Sinh viên có khả năng xử lý các công cụ phân tích dữ liệu tự động nhằm tìm ra các thông tin hữu ích, các đặc trưng tiềm ẩn ảnh hưởng để mục tiêu bài toán.
- **Thiết kế và cài đặt các thuật toán máy học:** Sinh viên được yêu cầu có khả năng đề xuất, triển khai và giải thích các thuật toán máy học nhằm giải quyết bài toán được giao.

1.3 Yêu cầu

Sinh viên phải thực hiện các yêu cầu tối thiểu sau:

1. Đọc và phân tích các đặc trưng trong 2 tập tin được cung cấp. Trình bày các thông tin hữu ích (insights) tác động đến chi phí y tế cá nhân.
2. Cài đặt các thuật toán máy học đã được học để dự đoán chi phí y tế cá nhân.
3. Báo cáo kết quả đạt được sau quá trình phân tích và cài đặt. Từ đó nhận xét về các tác nhân ảnh hưởng mạnh/yếu tới chi phí y tế cá nhân.

Lưu ý:

- Sinh viên có thể thực hiện thêm nhiều phân tích và thuật toán ngoài chương trình học. Dựa trên kết quả báo cáo và phân tích, sinh viên sẽ được tính điểm cộng tối đa 20% điểm của đề án này. Điểm tối đa của đề án là min(điểm đề án + điểm cộng, 10).
- Báo cáo cần có bảng đánh giá đóng góp và tự đánh giá của các thành viên trong nhóm.
- Các thuật toán phải được cài đặt bằng ngôn ngữ **Python** trên môi trường **Google Colab**.

2 Quy định nộp bài

Sinh viên cần nộp 1 file nén duy nhất có tên **<MSSV>.zip/rar** (ví dụ: 1912000.zip) trong đó chứa:

- Báo cáo: dưới dạng **pdf** chứa nội dung phân tích, các số liệu, dẫn chứng, và thuật toán nhóm đã cài đặt.
- Mã nguồn: tập tin **.py** hoặc **.ipynb**. Lưu ý mã nguồn phải có khả năng chạy được (runable) trên môi trường colab đã được quy định ở mục 1.3.