

# Thực hành mini-project

## -Hồi quy tuyến tính và Tiên đoán-

TS.Tô Đức Khánh

01/12/2024

Chọn 1 trong số các projects được trình bày dưới đây.

### Project 01 - Phát triển bán hàng trên app mobile hay website

Tiếp tin `Ecommerce_Customers.csv` có dữ liệu của khách hàng mua quần áo trực tuyến. Cửa hàng cung cấp các buổi tư vấn về phong cách và quần áo tại cửa hàng. Khách hàng đến cửa hàng, gặp gỡ/làm việc với nhà tạo mẫu cá nhân, sau đó họ có thể về nhà và đặt mua quần áo họ muốn trên ứng dụng di động hoặc trang web.

Công ty đang cố gắng quyết định xem nên tập trung nỗ lực vào trải nghiệm ứng dụng di động hay trang web của họ.

Dữ liệu bao gồm một số thông tin của khoảng 500 khách hàng:

- **Email:** địa chỉ email duy nhất xác định mỗi khách hàng.
- **Address:** địa chỉ của khách hàng.
- **Avg. Session Length:** trung bình thời gian tương tác.
- **Time on App:** thời gian tương tác trên mobile app.
- **Time on Website:** thời gian tương tác trên website.
- **Length of Membership:** thời gian là khách hàng thành viên.
- **Yearly Amount Spent:** chi phí mua hàng trung bình năm.

Hãy load dữ liệu và thực hiện một bản báo cáo cho nhà quản lý và phát triển của công ty, với các nội dung như sau:

1. Bảng tóm tắt, khái quát về dữ liệu.
2. Thời gian tương tác trên 2 nền tảng có khác biệt (có thể thử khai thác theo góc nhìn về thời gian là khách hàng thành viên)?
3. Lựa chọn biến  $Y$  để xây dựng mô hình cho thấy nên đầu tư vào nền tảng mobile app hay website.
4. Viết các nhận xét và kết luận.

### Project 02 - Lối sống và chỉ số khỏe mạnh

Tiếp tin `synthetic_health_data.csv` có dữ liệu của 1000 người, bao gồm:

- **Age:** tuổi của mỗi cá nhân (năm tuổi).
- **BMI:** chỉ số BMI của mỗi cá nhân.

- **Exercise\_Frequency:** số ngày tập thể dục trong 1 tuần.
- **Diet\_Quality:** chỉ số phản ánh chất lượng của chế độ ăn uống, với các giá trị cao hơn biểu thị thói quen ăn uống lành mạnh hơn.
- **Sleep\_Hours:** số giờ ngủ trung bình mỗi đêm.
- **Smoking\_Status:** biến nhị phân, trong đó 0 = không hút thuốc và 1 = hút thuốc.
- **Alcohol\_Consumption:** trung bình số lít chất có cồn tiêu thụ mỗi tuần.
- **Health\_Score:** điểm sức khỏe được tính toán phản ánh tình trạng sức khỏe tổng thể.

Trung tâm chăm sóc sức khỏe quốc gia muốn xây dựng một mô hình dự đoán điểm sức khỏe của công dân dựa trên độ tuổi và lối sống. Hãy load dữ liệu và thực hiện một bản báo cáo cho trung tâm chăm sóc sức khỏe quốc gia, với các nội dung như sau:

1. Bảng tóm tắt, khái quát về dữ liệu.
2. Xây dựng mô hình cho điểm sức khỏe.
3. Viết các nhận xét và kết luận.

## Project 03 - Xe ô tô và lượng phát thải khí CO2

Các tổ chức bảo vệ môi trường luôn quan tâm tới lượng phát thải khí CO2 của các phương tiện giao thông để đưa ra các khuyến cáo cho các chính phủ và các công ty nhằm thay đổi công nghệ sản xuất. Tiếp tin `vehicle_co2_emissions.csv` có dữ liệu của 7385 phương tiện cá nhân đang được sử dụng, bao gồm:

- **Brand:** thương hiệu hoặc nhà sản xuất của xe (ví dụ: Toyota, Ford, BMW).
- **Vehicle Type:** phân loại xe dựa trên kích thước và cách sử dụng (ví dụ: SUV, Sedan).
- **Engine Size (L):** thể tích dịch chuyển động cơ tính bằng lít.
- **Cylinders:** số xy-lanh trong động cơ.
- **Transmission:** loại hộp số (ví dụ: Tự động, Số sàn).
- **Fuel Type:** loại nhiên liệu mà xe sử dụng (D là dầu Diesel, E là Ethanol, N là compressed natural gas - khí thiên nhiên nén, X là regular gasoline - xăng, Z là premium gasoline - xăng cao cấp).
- **Fuel Consumption City (L/100 km):** mức tiêu thụ nhiên liệu của xe trong điều kiện lái xe trong thành phố, được đo bằng lít trên 100 km (L/100 km).
- **Fuel Consumption Hwy (L/100 km):** mức tiêu thụ nhiên liệu của xe trên đường cao tốc, được đo bằng lít trên 100 km (L/100 km).
- **Fuel Consumption Comb (L/100 km):** Mức tiêu thụ nhiên liệu kết hợp của xe (trung bình khi lái xe trong thành phố và trên đường cao tốc), được đo bằng lít trên 100 km (L/100 km).
- **Fuel Consumption Comb (mpg):** Mức tiêu thụ nhiên liệu kết hợp của xe được đo bằng số dặm (mile) trên 1 gallon nhiên liệu.
- **CO2 Emissions (g/km):** lượng khí thải carbon dioxide (gram) trên mỗi km

Hãy load dữ liệu và thực hiện một bản báo cáo cho tổ chức bảo vệ môi trường, với các nội dung như sau:

1. Bảng tóm tắt, khái quát về dữ liệu.
2. Xây dựng mô hình cho lượng khí thải carbon dioxide từ ô tô.
3. Viết các nhận xét và kết luận.