

# BÁO CÁO PROJECT 3

## 1. Thông tin cá nhân

- Họ và tên: Phan Trí Tài
- MSSV: 20127318

## 2. Mô tả hàm

### a) Thư viện

```
import pandas as pd
import numpy as np

# Import thêm dữ thư viện nếu cần
from sklearn import linear_model
from sklearn.metrics import mean_squared_error
import math
```

### b) Đọc dữ liệu

```
# Đọc dữ liệu bằng pandas
train = pd.read_csv('train.csv')
test = pd.read_csv('test.csv')

# Lấy các đặc trưng X và giá trị mục tiêu y cho các tập huấn luyện (train) và kiểm tra (test)
X_train = train.iloc[:, :-1] # Dataframe (chứa 10 đặc trưng huấn luyện)
y_train = train.iloc[:, -1]  # Series (chứa 1 giá trị mục tiêu kiểm tra)

X_test = test.iloc[:, :-1]    # Dataframe (chứa 10 đặc trưng kiểm tra)
y_test = test.iloc[:, -1]     # Series (chứa 1 giá trị mục tiêu kiểm tra)

# Sinh viên có thể sử dụng các khác nếu cần
x_data = train.iloc[:, :-1].to_numpy()
y_data = train.iloc[:, -1].to_numpy()
```

### c) Yêu cầu 1a

- Hàm đọc và in bảng dữ liệu của 'train.csv' gồm 1085 hàng  $\times$  11 cột

```
#Đọc dữ liệu
d_frame = pd.read_csv('train.csv')
d_frame
```

- Hàm dùng để chạy mô hình Linear Regression

```
#Chạy mô hình Linear Regression
l_regression = linear_model.LinearRegression().fit(properties, rate)
```

- Phần code cho yêu cầu 1a

```
# Phần code cho yêu cầu 1a
x = l_regression.coef_
y = l_regression.intercept_

print('Life expectancy = ',end = ' ')
for i in range(len(d_frame.columns) - 1):
    print('(x',end = '')
    print(i + 1,end = ')')
    print(' * ',end = '')
    print(l_regression.coef_[i],end = ' ')
    print('+',end = ' ')

print(l_regression.intercept_)
```

Kết quả:

```
Life expectancy = (x1) * -0.01667350187349808 + (x2) * 0.04572407012456567 + (x3) * 0.0014579648188069456 + (x4) * 0.025288029
555287975 + (x5) * -0.49803865291442306 + (x6) * 6.293928921841699e-05 + (x7) * -0.04033713679384392 + (x8) * -0.05648103715995
554 + (x9) * 12.705895973563242 + (x10) * 0.7076945141058034 + 52.433403436624864
```

- Gọi hàm RMSE (tự cài đặt hoặc từ thư viện) trên tập kiểm tra

```
# Gọi hàm RMSE (tự cài đặt hoặc từ thư viện) trên tập kiểm tra
p_1a = l_regression.predict([x_data[29]])
a_1a = y_data[29]

MSE = mean_squared_error([a_1a], [p_1a])

RMSE = math.sqrt(MSE)
print("Mean squared error: ",end = ' ')
print(RMSE)
```

Kết quả:

```
Mean squared error: 0.21833243601945185
```

- d) Yêu cầu 1b (chưa làm được)
- e) Yêu cầu 1c (chưa làm được)

### 3. Nhận xét

- Bài làm chỉ hoàn thành được 30% vì còn nhiều thiếu sót và vẫn chưa hoàn toàn hiểu được hết project.

### 4. Tài liệu tham khảo

- <https://github.com/t3bol90/ST-MA-Lab04/blob/master/18127231.ipynb>
- Tài liệu `project03.ipynb` trên Moodle
- Tài liệu `Lab 04 (Updated MSE function)File` trên Moodle