

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
ĐẠI HỌC QUỐC GIA – THÀNH PHỐ HỒ CHÍ MINH**

-----o0o-----



TRỰC QUAN HÓA DỮ LIỆU

REPORT

LAB 01

Nhóm sinh viên thực hiện:

20120041 – Trần Kim Bảo

20120053 – Nguyễn Thành Đạt

20120071 – Nguyễn Thị Bích Hà

20120084 – Nguyễn Văn Hiếu

20120113 – Lê Nguyên Khang

Thành phố Hồ Chí Minh, tháng 4 năm 2023

I. BẢNG THÀNH VIÊN NHÓM

MSSV	Họ và tên
20120041	Trần Kim Bảo
20120053	Nguyễn Thành Đạt
20120071	Nguyễn Thị Bích Hà
20120084	Nguyễn Văn Hiếu
20120113	Lê Nguyên Khang

II. BẢNG PHÂN CÔNG CÔNG VIỆC

Việc tìm kiếm bộ dữ liệu tất cả thành viên đều thực hiện tìm kiếm chung.

MSSV	Nội dung	Mức độ hoàn thành
20120041	<ul style="list-style-type: none">Mỗi dòng có ý nghĩa gì? Có vấn đề các dòng có ý nghĩa khác nhau không?Mỗi cột có ý nghĩa gì?Thực hiện phân tích: Khám phá mối quan hệ giữa các thuộc tính `Time ...` đến thuộc tính `Time utilized`.	100%
20120053	<ul style="list-style-type: none">Mỗi cột hiện đang có kiểu dữ liệu gì? Có cột nào có kiểu dữ liệu chưa phù hợp để có thể xử lý tiếp hay không?Thực hiện trả lời câu hỏi: Liệu có mối quan hệ nào giữa thuộc tính Time spent	100%

	on Online Class và Medium for online class đối với Rating of Online Class experience?	
20120071	<ul style="list-style-type: none"> Với mỗi cột, các giá trị (dạng số, dạng phân loại) được phân bố như thế nào? Phân tích tổng quan bộ dữ liệu (sự phân bố giá trị ở mỗi thuộc tính) Thực hiện trả lời câu hỏi: Liệu có mối quan hệ nào giữa hai cặp thuộc tính (dạng numeric) bất kì với nhau? Tổng hợp và đồng bộ hóa báo cáo (Merge tay file jupyter và kiểm tra lỗi trình bày, chính tả. 	100%
20120084	<ul style="list-style-type: none"> Thực hiện trả lời câu hỏi: Thiết bị nào được các bạn học sinh, sinh viên sử dụng để phục vụ cho việc học online nhiều nhất. Liệu có phải thời lượng các bạn học sinh, sinh viên tham học online phụ thuộc vào Thiết bị mà các bạn đó sử dụng? và Nền tảng mạng xã hội nào được sử dụng nhiều nhất theo độ tuổi? Kiểm tra lại file jupyter sau khi merge. 	100%
20120113	<ul style="list-style-type: none"> Có cần phải tiền xử lý dữ liệu hay không và nếu có thì nhóm sinh viên cần phải xử lý như thế nào? Thực hiện trả lời câu hỏi: Có phải những người mắc vấn đề sức khỏe trong quá 	100%

	trình đóng cửa xã hội, thường là những đối tượng có max hoặc min trong 1 thuộc tính hay tương quan với thuộc tính nào đó hay không?	
--	---	--

Link git của nhóm: <https://github.com/kuromine19/DataVisualization>

III. MỘT SỐ LƯU Ý KHI RUN FILE NOTEBOOK

Do sự khác nhau giữa các phiên bản thư viện, nhóm đã gặp trục trặc trong việc chạy các cell code bên dưới.

```
# Before sklearn 1.2
# from sklearn.metrics import plot_roc_curve

# From sklearn 1.2
from sklearn.metrics import RocCurveDisplay
```

```
In [54]: fig,(ax1,ax2)=plt.subplots(1,2,figsize=(18,8))
ax1.plot(false_positive,true_positive,'r',label="ROC")
ax1.plot(np.linspace(0,1,50),np.linspace(0,1,50),'b',linestyle='--',label="Random classifier")
ax1.set_xlabel("False Positive Rate")
ax1.set_ylabel("True Positive Rate")
ax1.legend()

# Use version before sklearn 1.2
plot_roc_curve(LogisticRegression().fit(X,y),X,y,ax=ax2)

# Use if sklearn version 1.2
RocCurveDisplay.from_estimator(LogisticRegression().fit(X,y),X,y,ax=ax2)

plt.legend()
plt.suptitle("ROC curve(Receiver operating characteristic)")
plt.show()
```

ROC curve(Receiver operating characteristic)

Cụ thể ở cell code thứ 54, nhóm có cho gọi một hàm với chức năng vẽ biểu đồ, do sự khác nhau giữa các phiên bản thư viện Scikit-learn nên nhóm đã quyết định sẽ sử dụng cả hai cách import hàm.

- Đối với phiên bản **sklearn dưới 1.2**, sẽ import hàm bằng cách bên dưới và gọi tên hàm như khung (1) trên hình.

```
from sklearn.metrics import plot_roc_curve
```

- Đối với phiên bản **sklearn từ 1.2 trở lên**, sẽ import hàm bằng cách bên dưới và gọi tên hàm như khung (2) trên hình.

```
from sklearn.metrics import RocCurveDisplay
```

Ngoài ra, đối với các thư viện cần được install mới có thể sử dụng, nhóm đã để sẵn câu lệnh install ở cell code đầu tiên để tiện cho việc run file notebook.

Import thư viện

```
In [1]: !pip install vega  
!pip install squarify  
!pip install altair
```