

# BÁO CÁO BÀI TẬP

Môn học: Phương pháp học máy trong An toàn thông tin

Tên chủ đề: Lab 1: Setting Up Your Machine Learning for Cybersecurity Arsenal

GVHD: Đoàn Minh Trung

## 1. THÔNG TIN CHUNG:

Lớp: NT522.N11.ATCL.2

Nhóm: N07

STT	Họ và tên	MSSV	Email
1	Nguyễn Tú Ngọc	20521665	20521665@gm.uit.edu.vn

## 2. NỘI DUNG THỰC HIỆN:

STT	Công việc	Kết quả tự đánh giá
1		100%
2		100%

Phần bên dưới của báo cáo này là tài liệu báo cáo chi tiết của nhóm thực hiện.

# BÁO CÁO CHI TIẾT

## 1. Phép cộng, trừ hai ma trận numpy:

```
1 import numpy as np
2 A = np.array([[4, 5, 6],
3              [7, 8, 9],
4              [10, 11, 12]])
5 B = np.array([[1, 2, 3],
6              [4, 5, 6],
7              [7, 8, 9]])
8 KQ = A + B
9 print("A + B = \n",KQ)
```

A + B =

```
[[ 5  7  9]
 [11 13 15]
 [17 19 21]]
```

## 2. Sử dụng Pandas xử lý các yêu cầu:

```
import pandas as pd
import numpy as np
from google.colab import files
uploaded = files.upload()

# Đọc CSV thành Dataframe và hiển thị
print("\n1. Đọc CSV thành Dataframe và hiển thị:")
import io
df = pd.read_csv(io.StringIO(uploaded['Dataset.csv'].decode('utf-8')))
print(df)

# Chuyển index mặc định thành giá trị cột id
print("\n2. Chuyển index mặc định thành cột ID:")
id = [0, 1, 2, 3, 4]
df_add = df.insert(0, 'ID', id)
df_index = df.set_index("ID", inplace = True)
print(df)

# Sắp xếp dữ liệu theo nhiều cột
print("\n3. Sắp xếp dữ liệu theo nhiều cột:")
df_sort = df.sort_values(["X", "Y"])
print(df_sort)

# Chọn 1 cột và hiển thị
print("\n4. Chọn 1 cột và hiển thị:")
df_column = df[["Y"]]
print(df_column)
```

```

# Hien thi 2 hang dau tien
print("\n5. Hiện thị 2 hàng đầu tiên:")
df_2row = df.head(2)
print(df_2row)

# Hien thi 1 hang dua tren dkien gtri 1 cot
print("\n6. Hiện thị 1 hàng dựa trên điều kiện giá trị của 1 cột:")
df_dkien = df[df['Z']>=86].head()
print(df_dkien)

# Thay doi gtri NaN bang gtri 0
print("\n7. Thay đổi giá trị NaN bằng giá trị 0:")
df_null = df.fillna('0')
print(df_null)

# Cot Z: Chuyen gtri > 90: True ; nguoc lai: False
print("\n8. Ở cột Z, chuyển giá trị > 90: True ; ngược lại: False")
df[["Z"]] = np.where(df[["Z"]]>90, True, False)
print(df)

# Join 2 dataframe
print("\n8. Join 2 dataframe thành 1:")

print("\nd1:")
data1 = df[["X" , "Y"]]
print(data1)

print("\nd2:")
data2 = df[["Z"]]
print(data2)

print("\nd3:")
data_join = pd.concat([data1, data2],axis=1)
print(data_join)

# Thong ke cac gtri thuoc tinh
print("\n9. Thống kê các giá trị thuộc tính:")
df.info()

```

### \*KẾT QUẢ:

```

1. Đọc CSV thành Dataframe và hiển thị:
      X      Y      Z
0   NaN  84.0  86.0
1  85.0  94.0  97.0
2   NaN  83.0   NaN
3  96.0  94.0  96.0
4  86.0   NaN  83.0

```

## Lab 01: Setting Up Machine Learning For Cybersecurity Arsenal

2. Chuyển index mặc định thành cột ID:

	X	Y	Z
ID			
0	NaN	84.0	86.0
1	85.0	94.0	97.0
2	NaN	83.0	NaN
3	96.0	94.0	96.0
4	86.0	NaN	83.0

3. Sắp xếp dữ liệu theo nhiều cột:

	X	Y	Z
ID			
1	85.0	94.0	97.0
4	86.0	NaN	83.0
3	96.0	94.0	96.0
2	NaN	83.0	NaN
0	NaN	84.0	86.0

4. Chọn 1 cột và hiển thị:

	Y
ID	
0	84.0
1	94.0
2	83.0
3	94.0
4	NaN

5. Hiển thị 2 hàng đầu tiên:

	X	Y	Z
ID			
0	NaN	84.0	86.0
1	85.0	94.0	97.0

6. Hiển thị 1 hàng dựa trên điều kiện giá trị của 1 cột:

	X	Y	Z
ID			
0	NaN	84.0	86.0
1	85.0	94.0	97.0
3	96.0	94.0	96.0

7. Thay đổi giá trị NaN bằng giá trị 0:

	X	Y	Z
ID			
0	0	84.0	86.0
1	85.0	94.0	97.0
2	0	83.0	0
3	96.0	94.0	96.0
4	86.0	0	83.0

8. Ở cột Z, chuyển giá trị > 90: True ; ngược lại: False

	X	Y	Z
ID			
0	NaN	84.0	False
1	85.0	94.0	True
2	NaN	83.0	False
3	96.0	94.0	True
4	86.0	NaN	False

## Lab 01: Setting Up Machine Learning For Cybersecurity Arsenal

9. Join 2 dataframe thành 1:

d1:

	X	Y
ID		
0	NaN	84.0
1	85.0	94.0
2	NaN	83.0
3	96.0	94.0
4	86.0	NaN

d2:

	Z
ID	
0	False
1	True
2	False
3	True
4	False

d3:

	X	Y	Z
ID			
0	NaN	84.0	False
1	85.0	94.0	True
2	NaN	83.0	False
3	96.0	94.0	True
4	86.0	NaN	False

10. Thống kê các giá trị thuộc tính:

<class 'pandas.core.frame.DataFrame'>

Int64Index: 5 entries, 0 to 4

Data columns (total 3 columns):

#	Column	Non-Null Count	Dtype
0	X	3 non-null	float64
1	Y	4 non-null	float64
2	Z	5 non-null	bool

dtypes: bool(1), float64(2)

memory usage: 125.0 bytes

### 3. Thực hiện mô hình Linear Regression bằng các thư viện: