

BÁO CÁO LAB

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

LAB 1: Setting Up Your Machine Learning for Cybersecurity Arsenal

Nhóm: 05

THÔNG TIN CHUNG:

Lớp: NT522.N21.ATCL

STT	Họ và tên	MSSV
1	Đỗ Quang Thắng	20521893
2	Nguyễn Đoàn Thiên Cung	20521146

1. NỘI DUNG THỰC HIỆN:

BÁO CÁO CHI TIẾT

Code bài Lab và dataset :

Link drive :

<https://colab.research.google.com/drive/1LIY1NM8gNWMHCyJ1JvIw9FGtCxLleNvs?usp=sharing>

Bài 5 : Sinh viên cho biết chức năng của phương thức `genfromtxt()` trong thư viện `numpy`.

- Phương thức **`genfromtxt()`** trong thư viện `numpy` được sử dụng để tải dữ liệu từ tệp văn bản (text file) vào mảng `numpy`. Đây là một trong những phương thức phổ biến nhất được sử dụng để xử lý dữ liệu trong khoa học dữ liệu và machine learning.
- Cú pháp của phương thức **`genfromtxt()`** như sau:

```
numpy.genfromtxt(fname, delimiter=',', skip_header=0, skip_footer=0,  
dtype=float, encoding=None, usecols=None,  
names=None, filling_values=None,  
comments=None, invalid_raise=True,
```

```
max_rows=None, encoding_errors='strict')
```

Trong đó:

- `fname`: tên tệp tin hoặc đường dẫn đến tệp tin chứa dữ liệu cần tải vào mảng numpy.
 - `delimiter`: ký tự phân tách các giá trị trong tệp tin, mặc định là dấu phẩy (,).
 - `skip_header`: số dòng bỏ qua tính từ đầu tệp tin.
 - `skip_footer`: số dòng bỏ qua tính từ cuối tệp tin.
 - `dtype`: kiểu dữ liệu của các phần tử trong mảng numpy, mặc định là float.
 - `encoding`: định dạng mã hóa của tệp tin, mặc định là None.
 - `usecols`: chỉ định các cột cần tải vào mảng, mặc định là tất cả các cột.
 - `names`: chỉ định tên cho các cột trong mảng numpy.
 - `filling_values`: giá trị được sử dụng để điền vào các ô trống trong tệp tin.
 - `comments`: ký tự bắt đầu dòng để xác định các dòng bị bỏ qua.
 - `invalid_raise`` nếu True, phương thức sẽ ném ra một ngoại lệ nếu có giá trị không hợp lệ trong tệp tin.
 - `max_rows`: chỉ định số dòng tối đa được tải vào mảng numpy.
 - `encoding_errors`: định dạng mã lỗi (nếu có) của tệp tin.
- Khi được gọi, phương thức **genfromtxt()** sẽ trả về một mảng numpy chứa các giá trị được tải từ tệp tin.

Ví dụ:

```
import numpy as np
```

```
data = np.genfromtxt('data.csv', delimiter=',', skip_header=1, dtype=float)
```

- Trong ví dụ này, phương thức `genfromtxt()` được sử dụng để tải dữ liệu từ tệp **data.csv** vào mảng numpy. Các giá trị được phân tách bằng dấu phẩy, và phương thức sẽ bỏ qua dòng đầu tiên của tệp tin. Kiểu dữ liệu của các phần tử trong mảng là float. Kết quả trả về là một mảng numpy chứa dữ liệu từ tệp tin.