

## **BÁO CÁO LAB**

Môn học: Phương pháp học máy trong an toàn thông tin LAB 2 : Machine Learning based Malware Detection

Nhóm: 05

#### **THÔNG TIN CHUNG:**

Lớp: NT522.N21.ATCL

STT	Họ và tên	MSSV	Phân công	Hoàn thành
1	Đỗ Quang Thắng	20521893	2,6,7,8	100%
2	Nguyễn Đoàn Thiên Cung	20521146	1,4,5	100%

#### 1. NỘI DUNG THỰC HIỆN:

# BÁO CÁO CHI TIẾT

Câu 1,4,5:

https://drive.google.com/drive/folders/13kQGk\_janXoDqHyBT07cnDAZIeQJXKHw?usp=sharing

Câu 2,6,7,8:

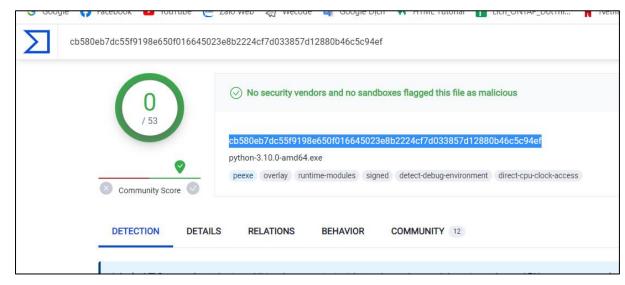
https://colab.research.google.com/drive/1egTPqvqVpKxrKmsBUZXJh1DceB9iK\_qk?usp=sharing



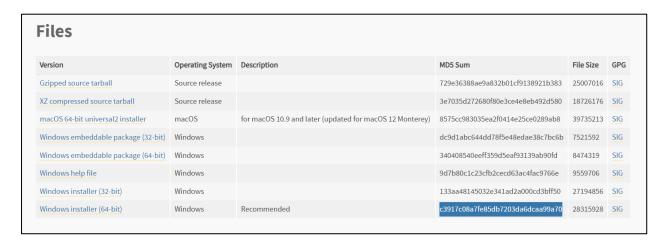
#### 1. Sinh viên so sánh kết quả băm với VirusTotal và website Python.

Hàm băm của code so với VirusTotal và website Python là giống nhau

```
Câu 1: Sinh viên so sánh kết quả băm với VirusTotal và website Python.
    import sys
     import hashlib
     filename = "/content/drive/MyDrive/Colab Notebooks/LAB2/python-3.10.0-amd64.exe"
     BUF_SIZE = 65536
    md5 = hashlib.md5()
     sha256 = hashlib.sha256()
    with open(filename, "rb") as f:
      while True:
         data = f.read(BUF_SIZE)
        if not data:
          break
        md5.update(data)
        sha256.update(data)
     print("MD5: {0}".format(md5.hexdigest()))
     print("SHA256: {0}".format(sha256.hexdigest()))
    MD5: c3917c08a7fe85db7203da6dcaa99a70
    SHA256: cb580eb7dc55f9198e650f016645023e8b2224cf7d033857d12880b46c5c94ef
```







2. Thực hiện đoạn code và in ra kết quả.

#### b) YARA

```
Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows
PS C:\Users\DELL\Downloads\New folder (6)> ls
     Directory: C:\Users\DELL\Downloads\New folder (6)
Mode
                             LastWriteTime
                                                            Length Name
                      4/6/2023
                                                           179 rules.yara
2221001 yara-4.3.0-2120-win64.zip
                                    10:13 AM
                                    10:10 AM
                      4/6/2023
                                                           2406912 yara64.exe
                     3/24/2023
                                      9:52 PM
                                                           2353664 yarac64.exe
PS C:\Users\DELL\Downloads\New folder (6)> .\rules.yara .\rules.yara '.\Lab 1 - Setting Up Your ML for Cybersecurity Ars
PS C:\Users\DELL\Downloads\New folder (6)> .\yara64.exe .\rules.yara '.\Lab 1 - Setting Up Your ML for Cybersecurity Ars
enal (1).pdf'
is_a_pdf .\Lab 1 - Setting Up Your ML for Cybersecurity Arsenal (1).pdf
dummy_rule2 .\Lab 1 - Setting Up Your ML for Cybersecurity Arsenal (1).pdf
PS C:\Users\DELL\Downloads\New folder (6)> .\yara64.exe .\rules.yara 'Lab 1 - Setting Up Your ML for Cybersecurity Arsenal (1).pdf
is_a_pdf Lab 1 - Setting Up Your ML for Cybersecurity Arsenal (1).pdf
dummy_rule2 Lab 1 - Setting Up Your ML for Cybersecurity Arsenal (1).pdf
PS C:\Users\DELL\Downloads\New folder (6)>
```

c) Kiểm tra PE header



```
print(pe.dump_into())
    0001CF70h HIGHLOW
    0001CF7Fh HIGHLOW
   0001CF89h HIGHLOW
    0001CFACh HIGHLOW
    0001CFB6h HIGHLOW
    0001CFF5h HIGHLOW
[IMAGE_BASE_RELOCATION]
0x81598 0x0 VirtualAddress:
                                               0x1D000
0x8159C 0x4 SizeOfBlock:
                                               0хС4
   0001D01Dh HIGHLOW
   0001D03Ah HIGHLOW
   0001D058h HIGHLOW
   0001D07Bh HIGHLOW
   0001D0C1h HIGHLOW
   0001D0FEh HIGHLOW
   0001D14Dh HIGHLOW
   0001D180h HIGHLOW
    0001D1A8h HIGHLOW
```

d) Featurizing the PE header

```
print(imports_corpus)
print(ms_sections)
print(section_names)

1005 Header magic not found.'
Unable to obtain imports from /content/drive/MyDrive/Mim 3 (2022-2023)/MK_2/Phurmg pháp học máy/Colab Notebooks/Benign PE Samples 6/appcmd.exe

1005 Header magic not found.'
Unable to obtain imports from /content/drive/MyDrive/Mim 3 (2022-2023)/MK_2/Phurmg pháp học máy/Colab Notebooks/Benign PE Samples 6/Appcmd.exe

1005 Header magic not found.'
Unable to obtain imports from /content/drive/MyDrive/Mim 3 (2022-2023)/MK_2/Phurmg pháp học máy/Colab Notebooks/Benign PE Samples 6/Adamuninstall.exe

1005 Header magic not found.'
Unable to obtain imports from /content/drive/MyDrive/Mim 3 (2022-2023)/MK_2/Phurmg pháp học máy/Colab Notebooks/Benign PE Samples 6/Adamuninstall.exe

1006 Header magic not found.'
Unable to obtain imports from /content/drive/MyDrive/Mim 3 (2022-2023)/MK_2/Phurmg pháp học máy/Colab Notebooks/Benign PE Samples 6/AppvStreamIngUX.exe

1006 Header magic not found.'
Unable to obtain imports from /content/drive/MyDrive/Mim 3 (2022-2023)/MK_2/Phurmg pháp học máy/Colab Notebooks/Benign PE Samples 6/AppvStreamIngUX.exe

1007 Header magic not found.'
Unable to obtain imports from /content/drive/MyDrive/Mim 3 (2022-2023)/MK_2/Phurmg pháp học máy/Colab Notebooks/Malicious PE Samples 6/AppvStreamIngUX.exe

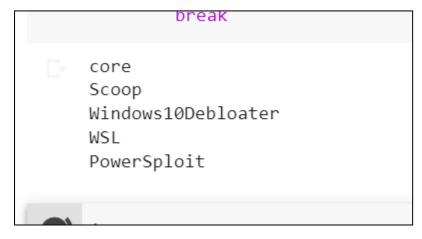
1008 Header magic not found.'
Unable to obtain imports from /content/drive/MyDrive/Mim 3 (2022-2023)/MK_2/Phurmg pháp học máy/Colab Notebooks/Malicious PE Samples 6/AppvStreamIngUX.exe

1008 Header magic not found.'
Unable to obtain imports from /content/drive/MyDrive/Mim 3 (2022-2023)/MK_2/Phurmg pháp học máy/Colab Notebooks/Malicious PE Samples 6/AppvStreamIngUX.exe

1007 Header magic not found.'
Unable to obtain imports from /content/drive/MyDrive/Mim 3 (2022-2023)/MK_2/Phurmg pháp học máy/Colab Notebooks/Malicious PE Samples 6/AppvStreamIngUX.exe

1008 Header magic not found.'
Unable to obtain imports from /content/drive/MyDrive/Mim 3 (2022-2023)/MK_2/Phurmg pháp học máy/Colab Notebooks/Malicious PE Samples
```

- 4. Tương tự sinh viên hãy làm các câu truy vấn về Python và Powershell
  - Powershell



- Python



```
public-apis
youtube-dl
core
awesome-machine-learning
ansible
```

#### 5. Sinh viên cho biết quả của đoạn code trên

- Kết quả ta thu được từ đoạn code đã cho

```
0.9732142857142857
[[122 0 0]
[ 0 82 0]
[ 6 0 14]]
```

## 6. Thực thi và kiểm tra kết quả.

g) Đo lường sự giống nhau giữa hai chuỗi

```
3:f4oo8MRwRJFGW1gC6uWv6MQ2MFS1+JuBF8BSnJi:f4kPvtHMCMubyFtQ
3:f4oo8MRwRJFGW1gC6uWv6MQ2MFS1+JuBF8BS+EFECJi:f4kPvtHMCMubyFIsJQ
3:f4oo8MRwRJFGW1gC6uWv6MQ2MFS1+JuBF8BS6:f4kPvtHMCMubyF0
3:60QKZ+4CDTfDaRFKYLVL:ywKDC2mVL
100
39
37
```

h) Đo lường mức độ giống nhau giữa hai tập tin



```
quangthang@quangthang-VirtualBox: $ hexdump -C python-3.10.0-amd64.exe | tail -5
01b010e0 10 9c 34 66 02 d3 51 8c b1 64 19 f3 55 12 0e 74 | ..4f..Q..d..U..t|
01b010f0 38 71 4c 2e 1c db 44 d4 f3 81 31 a5 9c 2e c6 06 | 8qL...D...1....|
01b01100 4f 33 c6 8a 9a 5e 16 52 8c 4b 55 10 2b cd 45 61 | 03...^R.KU.+.Ea|
01b01110 a5 00 00 00 00 00 00 00 00 | ......|
01b01118
quangthang@quangthang-VirtualBox: $ hexdump -C python-3.10.0-amd64-fake.exe | tail -5
01b010e0 10 9c 34 66 02 d3 51 8c b1 64 19 f3 55 12 0e 74 | ..4f..Q..d..U..t|
01b01100 4f 33 c6 8a 9a 5e 16 52 8c 4b 55 10 2b cd 60 | 8qL...D...1....|
01b01110 a5 00 00 00 00 00 00 00 00 00 | ......|
01b011110 a5 00 00 00 00 00 00 00 00 00 | ......|
01b01118
quangthang@quangthang-VirtualBox: $ python3 compare.py
100
quangthang@quangthang-VirtualBox: $ python3 compare.py
```

- 7. Thực thi và kiểm tra kết quả.
- i) Trích xuất N-grams

```
print(extracted_Ngrams.most_common(10))

[((0, 0, 0, 0), 24290), ((139, 240, 133, 246), 1920), ((32, 116, 111, 32), 1791), ((255, 255, 255), 1671), ((108, 101, 100, 32), 1522), ((100, 32, 116, 111), 1519), ((97, 105), (100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 100, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 101, 10
```

j) Chọn N-grams tốt nhất



```
print("Frequency:")
print(X, Top, X2, Freq)
print("Kinds Information:")
print(X, Top, X2, Ini)
print("X, Top, X2, Ini)
print("
```

### 8. Thực thi và kiểm tra kết quả.

```
print("Classifier score: \n", classifier.score(X_test, y_test))

/content/drive/MyOrive/Nām 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Benign PE Samples 6/aspnetca.exe:
'DOS Header magic not found.'
/content/drive/MyOrive/Nām 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Benign PE Samples 6/AppVStreamingUX.exe:
'PE' object has no attribute 'DIRECTORY_ENTRY_IMPORT'
/content/drive/MyOrive/Nām 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Malicious PE Samples 2/Build.exe:
'utf-8' codec can't decode byte 0xd2 in position 6: invalid continuation byte
/content/drive/MyOrive/Nām 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Benign PE Samples 6/ADSchemaAnalyzer.exe:
'DOS Header magic not found.'
/content/drive/MyOrive/Nām 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Benign PE Samples 6/adamuninstall.exe:
'DOS Header magic not found.'
/content/drive/MyOrive/Nām 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Benign PE Samples 6/appcmd.exe:
'DOS Header magic not found.'
/content/drive/MyOrive/Nām 3 (2022-2023)/HK_2/Phương pháp học máy/Colab Notebooks/Benign PE Samples 6/appcmd.exe:
'DOS Header magic not found.'
Classifier score:
0.8571428571428571
```

