المعهد الوطني للبريد والمواصلات

Institut National des Postes et Télécommunications

Using machine learning to predict phishing sites

La base de sécurité

DE

Réalisé par :

Youssef JEHBALI

Mohamed Khalil MOUADI

El Hassan CHAKOUKI

ABID Taib

Encadré par :

Mme Meryeme AYACHE

Sommaire

I. Introduction

1) *Definition*

2) *Phishing attack examples*

II. What’s machine learning ?

1) *Definition*

2) *Machine Learning Methods*

3) *Some Machine Learning Techniques*

III. Using machine learning to predict phishing sites

1) *Importing some useful libraries*

2) *Loading the main dataset*

3) *Preprocessing*

4) *LogisticRegression*

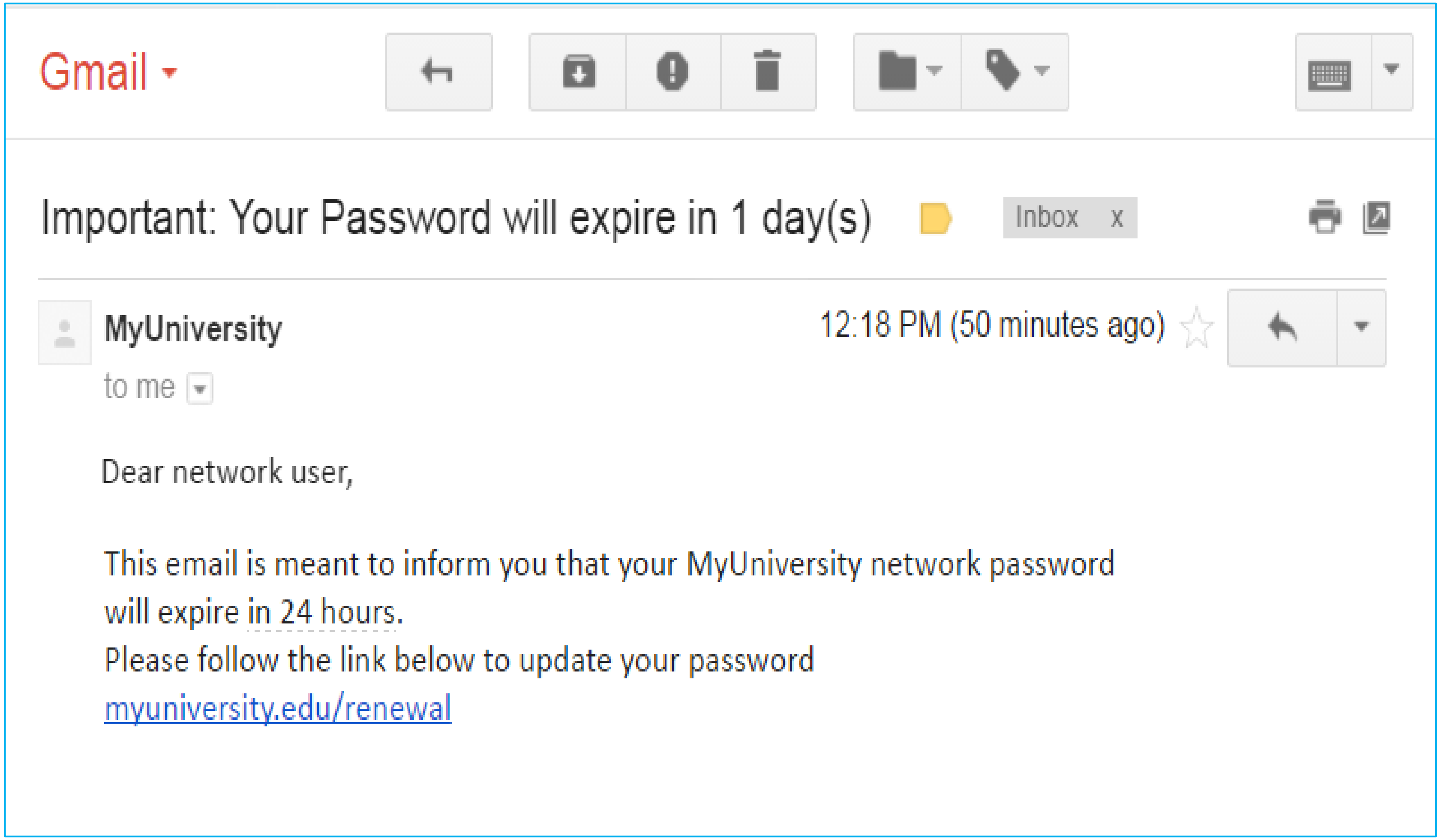
IV. Conclusion

I. Introduction

1- Définition :

Phishing est un type d'attaque d'ingénierie sociale souvent utilisée pour voler les données des utilisateurs, y compris les identifiants de connexion et les numéros de carte de crédit. Cela se produit lorsqu'un attaquant, se faisant passer pour une entité de confiance, dupe une victime pour qu'elle ouvre un e-mail, un message instantané ou un message texte. Le destinataire est alors amené à cliquer sur un lien malveillant, ce qui peut conduire à l'installation de logiciels malveillants, au blocage du système dans le cadre d'une attaque de ransomware ou à la révélation d'informations sensibles.

2- Example of phishing attack :



Important: Your Password will expire in 1 day(s)

MyUniversity

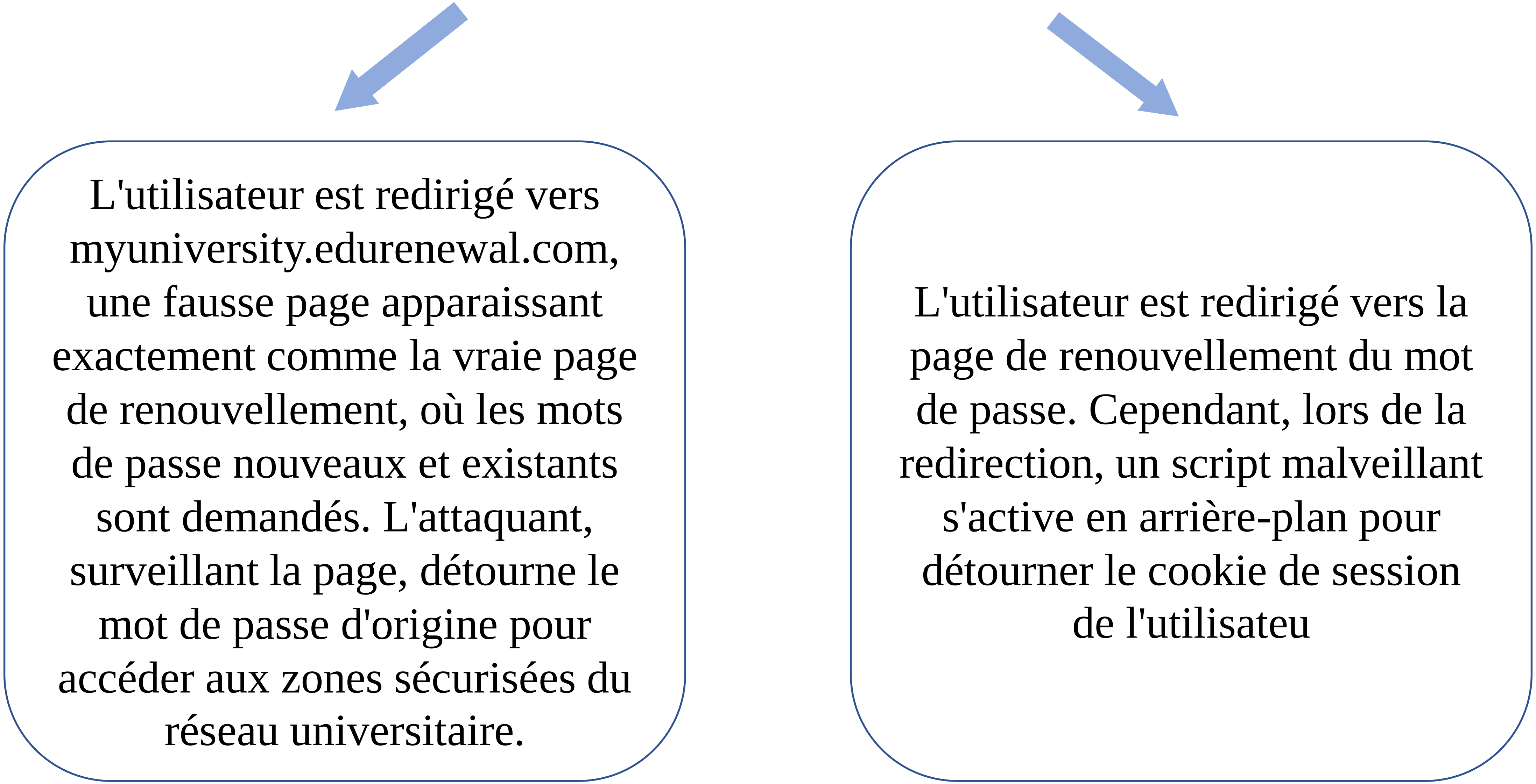
to me

12:18 PM (50 minutes ago)

Dear network user,

This email is meant to inform you that your MyUniversity network password will expire in 24 hours.
Please follow the link below to update your password
myuniversity.edu/renewal

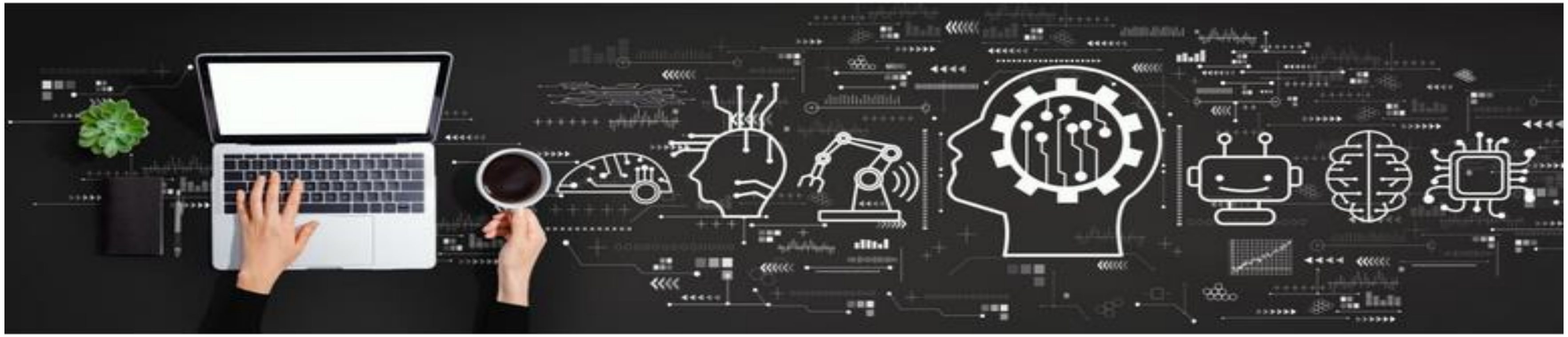
Plusieurs choses peuvent se produire en cliquant sur le lien



L'utilisateur est redirigé vers myuniversity.edurenewal.com, une fausse page apparaissant exactement comme la vraie page de renouvellement, où les mots de passe nouveaux et existants sont demandés. L'attaquant, surveillant la page, détourne le mot de passe d'origine pour accéder aux zones sécurisées du réseau universitaire.

L'utilisateur est redirigé vers la page de renouvellement du mot de passe. Cependant, lors de la redirection, un script malveillant s'active en arrière-plan pour détourner le cookie de session de l'utilisateu

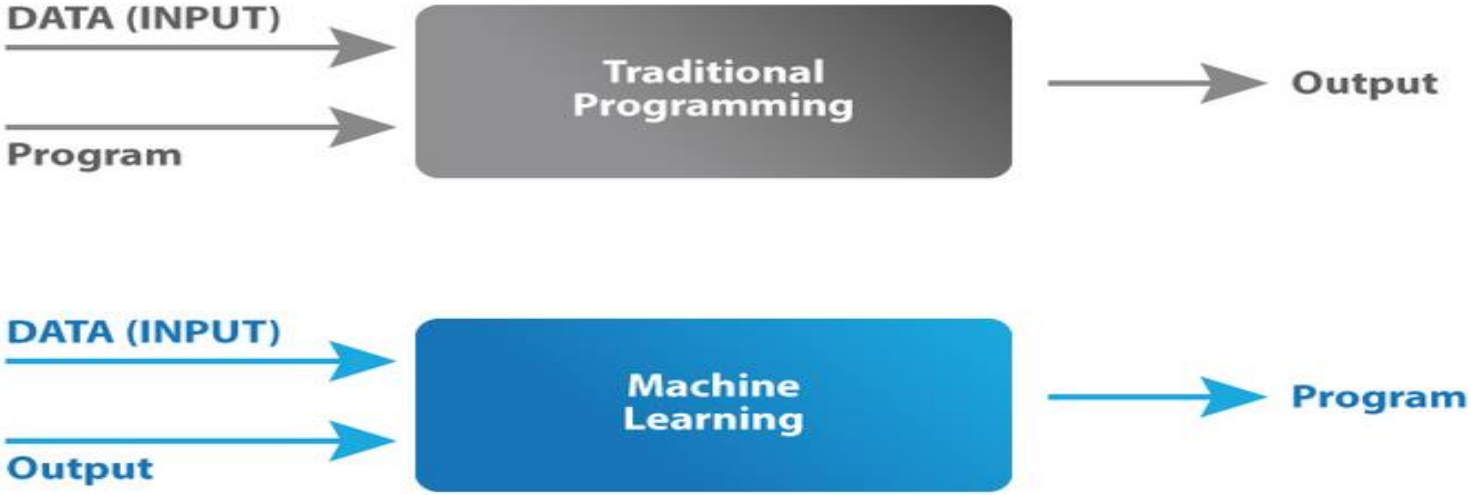
II.What’s machine learning ?!



1- Definition

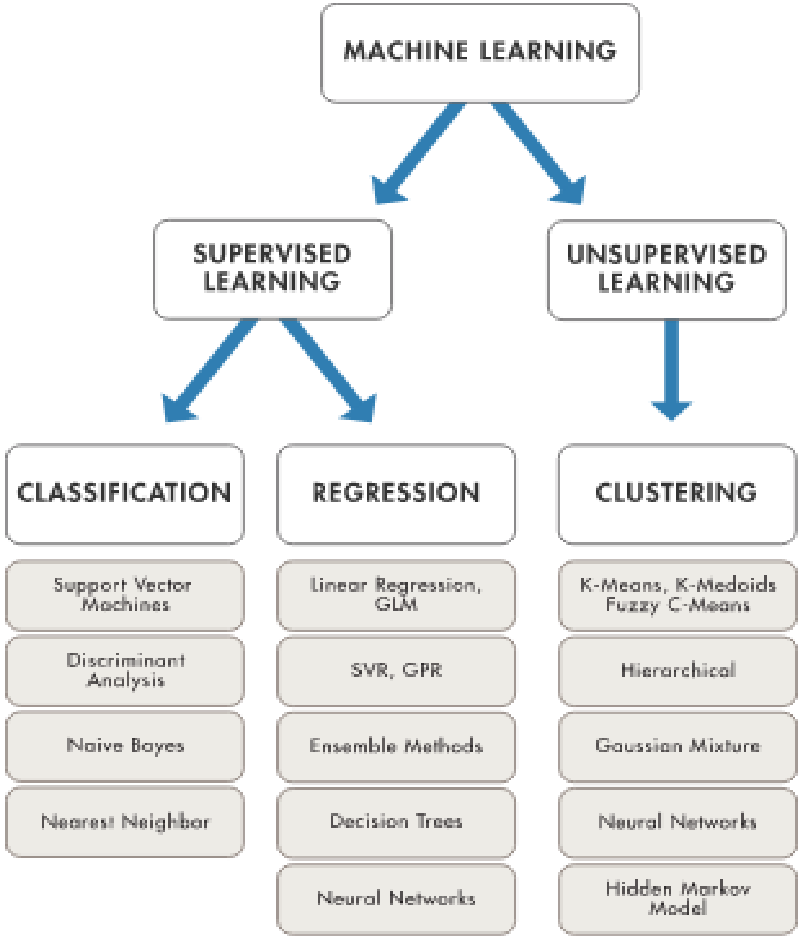
ML is an Application of Artificial Intelligence (AI) it gives devices the ability to learn from their experiences and improve their self without doing any coding. For Example, when you shop from any website it's shows related search like:

- People who bought also saw this.



2-Machine Learning Methods

Most ML algorithms fall into one of these categories:



When ?

We apply supervised ML techniques when we have a piece of data that we want to predict or explain. We do so by using previous data of inputs and outputs to predict an output based on a new input. By contrast, unsupervised ML looks at ways to relate and group data points without the use of a target variable to predict. In other words, it evaluates data in terms of traits and uses the traits to form clusters of items that are similar to one another.

3 -Some Machine Learning Techniques

ML is a hot topic in research and industry, with new methodologies developed all the time.

Essential ML Techniques :

a) Regression

Regression methods fall within the category of supervised ML. They help to predict or explain a particular numerical value based on a set of prior data, for example predicting the price of a property based on previous pricing data for similar properties.

The simplest method is linear regression where we use the mathematical equation of the line ($y = m * x + b$) to model a data set. We train a linear regression model with many data pairs (x, y) by calculating the position and slope of a line that minimizes the total distance between all of the data points and the line. In other words, we calculate the slope (m) and the y-intercept (b) for a line that best approximates the observations in the data.

b) Classification

Another class of supervised ML, classification methods predict or explain a class value. For example, they can help predict whether or not an online customer will buy a product. The output can be yes or no: buyer or not buyer

The simplest classification algorithm is logistic regression (which makes it sounds like a regression method, but it’s not). Logistic regression estimates the probability of an occurrence of an event based on one or more inputs

c) Clustering

With clustering methods, we get into the category of unsupervised ML because their goal is to group or cluster observations that have similar characteristics. Clustering methods don’t use output information for training, but instead let the algorithm define the output. In clustering methods, we can only use visualizations to inspect the quality of the solution.

The most popular clustering method is K-Means, where “K” represents the number of clusters that the user chooses to create.

III. Using machine learning to predict phishing sites

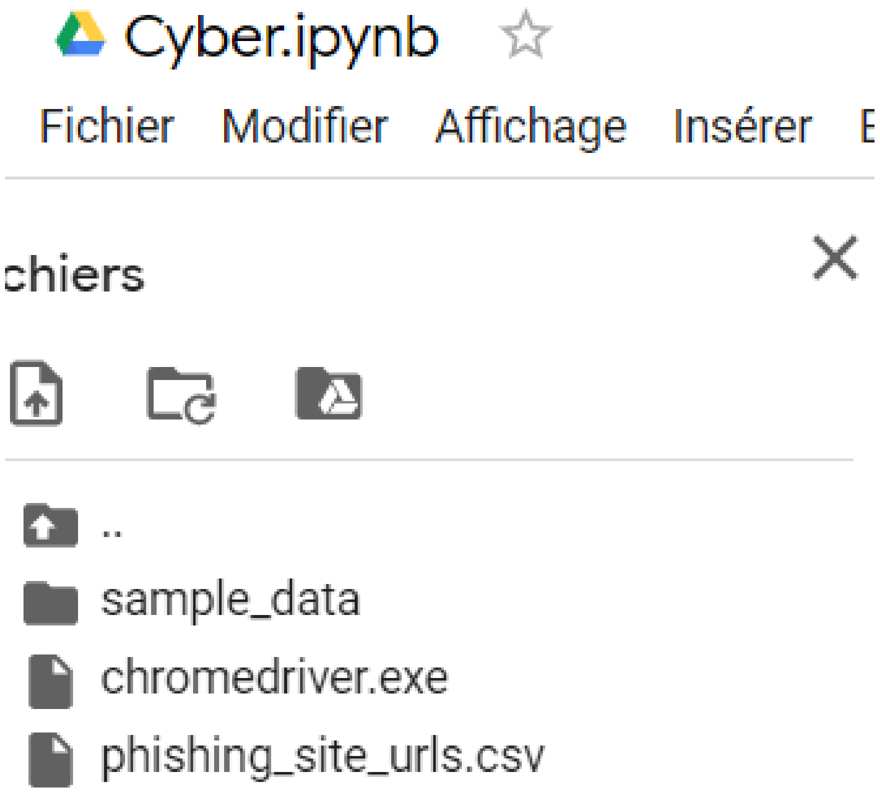
https://colab.research.google.com/drive/1-tHjJrWe6o-Qkj5DoexwBd7kexJDgmy7?usp=sharing

1) Importing some useful libraries

2) Loading the main dataset

3) Preprocessing

4) LogisticRegression



Cyber.ipynb
Fichier
Modifier
Affichage
Insérer
E
chiers
..
sample_data
chromedriver.exe
phishing_site_urls.csv

IV. Conclusion

THANK YOU