



SECURITÉ IT & CONFIANCE NUMERIQUE

Module :

M243 - Intelligence Artificielle et applications à la cybersécurité

TP1 — Exploration des données cybersécurité

Réalisé par :

Badra Aliou Keita

Activité 1 : compter les attaques



```
1 # Compter les attaques
2 from script import df
3 nb_attaques = (df["label"] == 1).sum()
4 print("Nombre d'attaques : ", nb_attaques)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER

```
└─(.env)─(draken㉿draken)─[~/Desktop/TP/IA/TP_1]
$ python3 activite_1.py
Nombre d'attaques : 162
```

Dans cette activité il s'agit de chercher le nombre total d'attaque introduit dans le DataSet.

Processus : étant donné que les attaque ont pour label 1, nous allons filtré par label en effectuant la somme.

Activité 2 : afficher uniquement les attaques

```
1 # Affichage des attaques
2 from script import df
3 attaques = df[df["label"] == 1]
4 print("Affichage des lignes correspondant à des attaques")
5 print(attaques)
6
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER

```
└─(.env)─(draken㉿draken)─[~/Desktop/TP/IA/TP_1]
$ python3 activite_2.py
Affichage des lignes correspondant à des attaques :
   feature_0  feature_1 ...  feature_11  label
17     -0.858566   0.871419 ...   1.144051      1
32      0.556876   0.101229 ...  -0.139858      1
50     -1.866342   1.832437 ...   2.186258      1
68     -0.477971  -0.788062 ...  -0.325312      1
76     -0.680938  -0.128233 ...   0.595569      1
...
2924    0.240773  -0.935314 ...   0.352154      1
2932    1.782491  -0.065344 ...   0.120949      1
2947    -3.744586   0.695109 ...   0.990287      1
2971    -0.316825   1.893498 ...   0.843744      1
2980    -0.897714   1.462990 ...  -0.002649      1
```

Dans cette partie nous allons procéder à la même manier que dans la première activité en filtrant par label.

Activité 3 : trouver la feature la plus variable



```
1 from script import df
2
3 print("\nActivité 3 :")
4 print("Futuré la plus variable ")
5
6 stds = df.drop(columns=["label"]).std()
7 feature_trace = stds.idxmax()
8 print(feature_trace)
9 print("Ecart-type : ", stds.max())
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER +

```
└─(.env)─(draken㉿draken) [~/Desktop/TP/IA/TP_1]
$ python3 activite_3.py
```

```
Activité 3 :
Futuré la plus variable
feature_3
Ecart-type : 2.16493024150287
```

Le feature dont les valeurs changent le plus d'un échantillon à l'autre.

Méthode de calcul :

Écart-type et la variance.

Activité 4 : tracer la distribution d'une feature

المدرسة الوطنية العليا للذكاء الاصطناعي وعلوم المعطيات - تارودانت
ÉCOLE NATIONALE SUPÉRIEURE DE L'INTELLIGENCE ARTIFICIELLE
ET SCIENCES DES DONNÉES - TAROUDANT



```
1 ~/Desktop/TP/IA/TP_1/activite_4.py
2 from activite_3 import feature_trace
3 import matplotlib.pyplot as plt
4
5 print("\nActivité 4 :")
6 # Nous allons tracer le feature trouve dans l'activité 3
7 plt.figure()
8 plt.hist(df[feature_trace], bins=30, color="skyblue", edgecolor="black")
9 plt.title(f"Distribution de {feature_trace}")
10 plt.xlabel("Valeur")
11 plt.ylabel("Fréquence")
12 plt.show()
13 plt.savefig("activite_4.png")
14
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER +

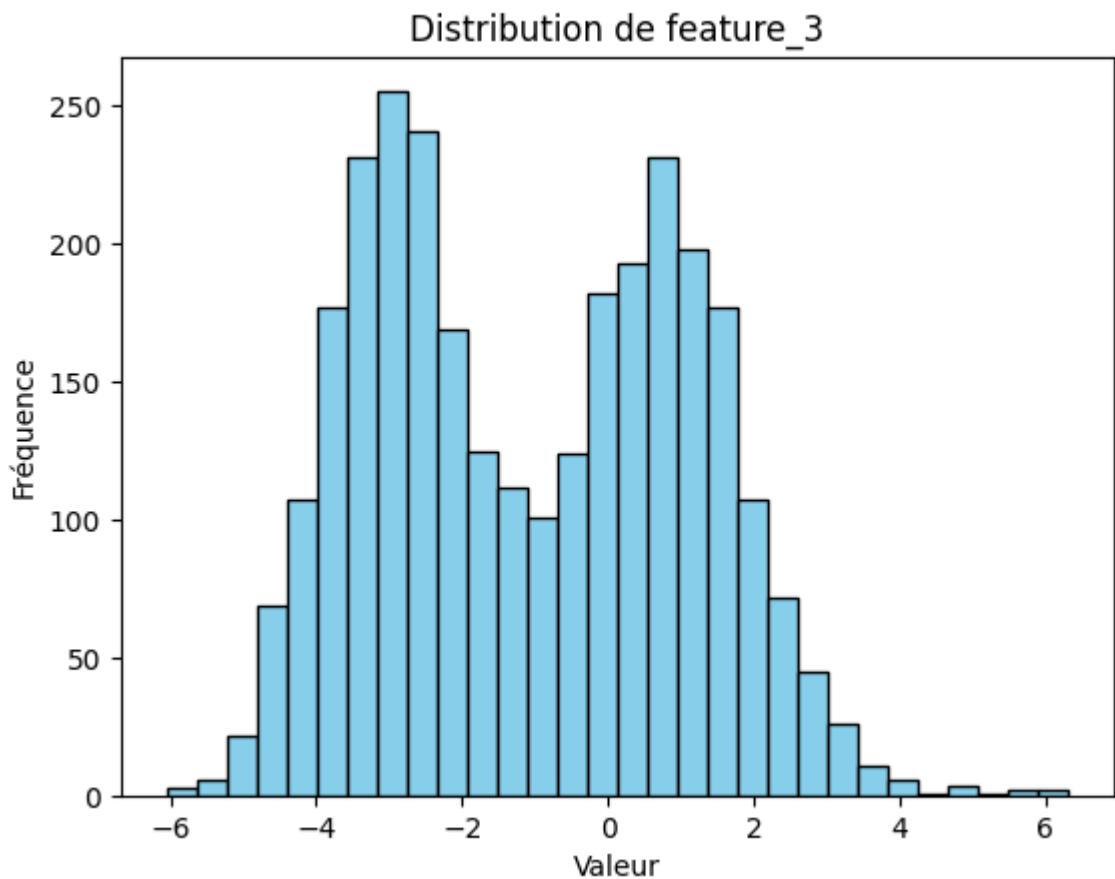
● (.) (.env)-draken@draken-[~/Desktop/TP/IA/TP_1]
\$ python3 activite_4.py

Activité 3 :

Futuré la plus variable
feature_3
Ecart-type : 2.16493024150287

Activité 4 :

Lors du tracer je vais utiliser la valeur de l'activité 3



La distribution de feature_3 montre comment ses valeurs se répartissent dans le dataset.

- On observe une **forte concentration** des valeurs entre -4 et 2.
- La fréquence maximale est d'environ **250 échantillons** pour la plage -4 à -2.
- Cette concentration indique que la majorité des observations se situe dans cette plage, tandis que les valeurs extrêmes sont rares.

Cela peut aider à comprendre quelles plages de valeurs sont typiques pour le trafic normal et quelles valeurs pourraient indiquer une attaque.

Réponse aux questions :



1. Les attaques sont rares car dans un réseau réel la majorité des requêtes sont normales, les attaquants essaient de passer inaperçue pour éviter la détection.
2. Le déséquilibre des données pose un problème car les modèles ont tendance à prédire la classe majoritaire.
3. Non, Les attaques se déroulent en plusieurs phases et les attaquants ont tendance à modifier les patterns connus.
4. Si les données sont bruitées, on peut nettoyer, supprimer ou corriger les valeurs aberrantes en utilisant des techniques de prétraitement de données appropriées.