

Challenge : Détection de flux réseau

Prêts à relever le défi ? Nous vous proposons un concours de modélisation autour de l'analyse de flux réseau !

Vous disposez d'un jeu de données d'entraînement intitulé `Training.parquet`, contenant des statistiques issues de flux réseau réels. Le champ `ClassLabel` indique la nature de chaque flux :

```
print(pd.unique(df["ClassLabel"]))

['Benign' 'Botnet' 'Bruteforce' 'DDoS' 'DoS' 'Infiltration'
 'Portscan' 'Webattack']
```

▣ Votre mission :

Entraîner un modèle de classification encapsulé dans un **pipeline scikit learn** afin qu'il puisse être chargé et évalué automatiquement par nos soins. L'évaluation sera effectuée sur le jeu de test `Testing.parquet` via le code suivant :

```
import pandas as pd
df = pd.read_parquet("Testing.parquet")
import skops.io as skio

from_model = None
with open("student_model.skio", "rb") as f:
    from_model = skio.load(f)

X = df.drop(columns=["ClassLabel"], axis=1)
y = df["ClassLabel"]

score = from_model.score(X, y)
```

▣ Classement et évaluation :

Chaque groupe sera classé selon le score obtenu. Une partie de la note finale sera pondérée par ce classement, donc **optimisez vos performances** !

▣ À rendre :

Le code complet de votre projet dans le dépôt Git qui vous sera attribué.

Le fichier `student_model.skio` à la racine de ce dépôt.

▣ À vous de jouer !

Faites preuve de créativité, d'efficacité et de rigueur pour construire le modèle le plus performant. Le meilleur pipeline remportera la première place du classement !