**Test technique :**

Vous trouverez ci-dessous un petit challenge technique. Il n’est pas demandé d’y consacrer de longues heures, mais il nous permet d’avoir une première vision de vos connaissances actuelles et de votre manière de raisonner.

Si le challenge se rapproche de ceux que vous pouvez trouver sur les plateformes de compétitions de machine-learning, l’objectif est plus large. Ainsi, si le résultat global sera regardé, ce n’est pas la partie la plus importante, nous regarderons aussi:

* La qualité des réponses aux questions.
* La qualité du code.
* La documentation et l’explication des choix.

**Le challenge :**

L'objectif de ce challenge est de prédire si une espèce de champignons est comestible ou bien vénéneuse. Cependant, comme aucune règle a priori ne permet de déduire la venimosité d'un champignon, un objectif secondaire mais tout aussi important est d'identifier les caractéristiques principales (importance de feature) pour déterminer si oui ou non un champignon sera comestible.

Nous attendons donc de vous :

* La construction d’un modèle de prédiction de la dangerosité (colonne « class »).
* Identifier et justifier les variables les plus pertinentes dans le modèles.
* Une courte explication sur la qualité de votre modèle et comment vous évaluer cette qualité.

**Les données :**

Pour ce faire, vous disposez d’un jeu de données avec des informations catégorielles sur chacun des aspects de différents types de champignons. 22 colonnes recensent ces informations et une autre donne la classe (e = comestible, p = vénéneux), qui sera la colonne à prédire. Ces informations sont données dans le fichier features.txt ci-joint.

Vous trouverez donc les fichiers suivants :

* Features.txt : qui donne des détails sur les variables.
* Train.csv : qui contient les données d’entrainements.

**Rendu :**

**Vous pouvez utiliser le langage de votre choix (R, Python, scala, Java, …).**

Merci de fournir :

* Le code
* Un petit fichier texte pour répondre aux questions ci-dessus.

Nous conservons une partie des données et évaluerons votre modèle sur la partie que nous conservons. C’est pourquoi nous attendons une structure de code précise pour tester votre implémentation.

Vous pouvez trouver ci-dessous un exemple en python de comment nous évaluerons votre modèle et les attendus en terme de réalisation. **Vous devez implémenter la fonction « predict ».** Bien entendu n’hésitez pas à poser des questions si ce n’est pas clair, ce ne sera en aucun cas pénalisant.

import pandas as pd

def predict(df):

  # TODO – doit renvoyer une Series pandas avec les valeurs ‘e’ pour comestible et ‘p’ pour vénéneux.

if \_\_name\_\_ == "\_\_main\_\_":

  df = pd.read\_csv('evaluation.csv') # Chargement de notre fichier csv d’évaluation : format équivalent à train.csv

  y = predict(df)

  evaluate(y, true\_y) # Notre fonction d’évaluation : vous n’avez pas à l’implémenter.