

## Brief TEST DE PERSONNALITÉ

### 1) Données, Analyse, Prétraitement.

Après avoir concaténer les fichiers importés en un seul dataframe(combined\_csv).  
La variable à prédire est interpretation des scores pour le niveau de stress.

Score :

- 1 ou a ou A vaut 1 point
- 2 ou b ou B vaut 0 point
- 3 ou c ou C vaut 2 points

Interpretation :

- score < 10 => C
- score < 20 => B
- score < 30 => A

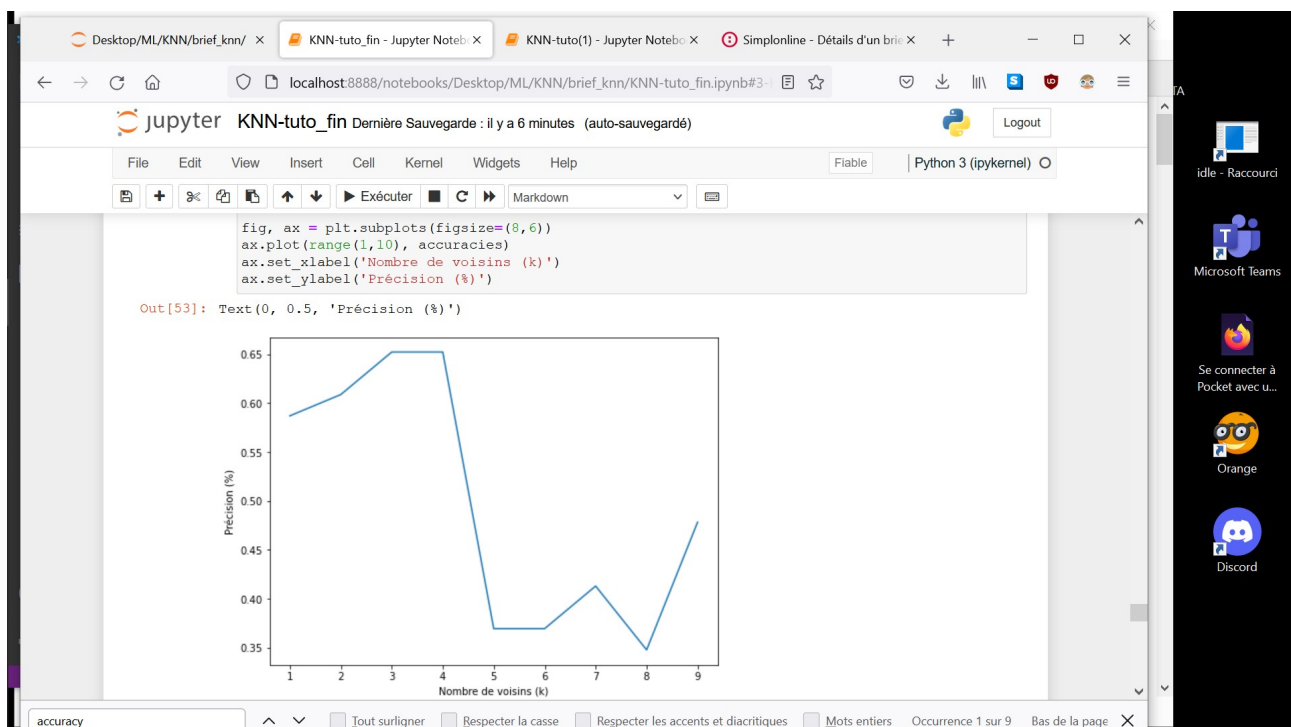
### 2) Développement et entraînement d'un modèle KNN from Scratch

2.1) définir une fonction capable de calculer la distance.

2.2) Classer ces distances par ordre croissant et récupérer les X plus proches voisins.

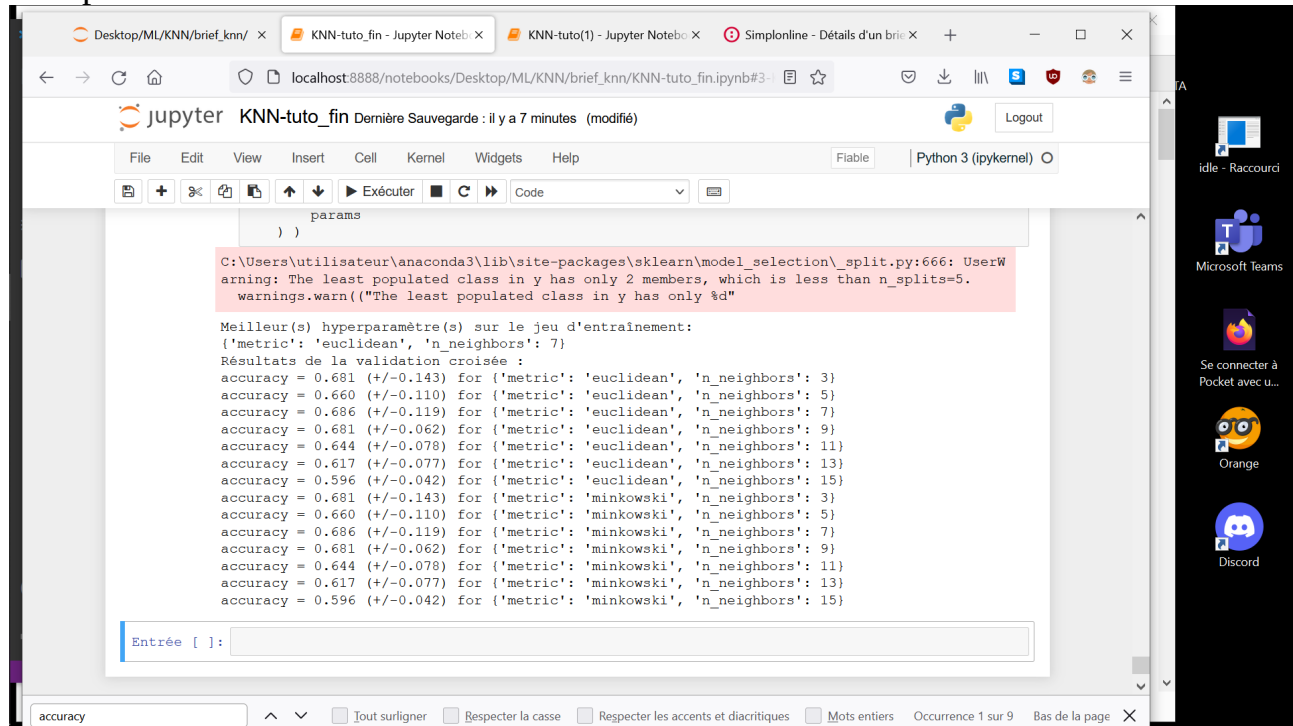
2.3) Trouver quel label est le plus commun parmi les voisins.

Résultat de modélisation : meilleur modèle concernant k=3 ou 4. qui correspond à la plus grande accuracy(0.67) ce qui est proche du modèle avec **Sklearn** (accuracy = 0.681 (+/- 0.143) for {'metric': 'euclidean', 'n\_neighbors': 3})



### 3) KNN Sklearn

Comparaison entre les modèles Sklearn et from scratch.



```
params
) )

C:\Users\utilisateur\anaconda3\lib\site-packages\sklearn\model_selection\_split.py:666: UserWarning: The least populated class in y has only 2 members, which is less than n_splits=5.
  warnings.warn("The least populated class in y has only %d"

Meilleur(s) hyperparamètre(s) sur le jeu d'entraînement:
{'metric': 'euclidean', 'n_neighbors': 7}
Résultats de la validation croisée :
accuracy = 0.681 (+/-0.143) for {'metric': 'euclidean', 'n_neighbors': 3}
accuracy = 0.660 (+/-0.110) for {'metric': 'euclidean', 'n_neighbors': 5}
accuracy = 0.686 (+/-0.119) for {'metric': 'euclidean', 'n_neighbors': 7}
accuracy = 0.681 (+/-0.062) for {'metric': 'euclidean', 'n_neighbors': 9}
accuracy = 0.644 (+/-0.078) for {'metric': 'euclidean', 'n_neighbors': 11}
accuracy = 0.617 (+/-0.077) for {'metric': 'euclidean', 'n_neighbors': 13}
accuracy = 0.596 (+/-0.042) for {'metric': 'euclidean', 'n_neighbors': 15}
accuracy = 0.681 (+/-0.143) for {'metric': 'minkowski', 'n_neighbors': 3}
accuracy = 0.660 (+/-0.110) for {'metric': 'minkowski', 'n_neighbors': 5}
accuracy = 0.686 (+/-0.119) for {'metric': 'minkowski', 'n_neighbors': 7}
accuracy = 0.681 (+/-0.062) for {'metric': 'minkowski', 'n_neighbors': 9}
accuracy = 0.644 (+/-0.078) for {'metric': 'minkowski', 'n_neighbors': 11}
accuracy = 0.617 (+/-0.077) for {'metric': 'minkowski', 'n_neighbors': 13}
accuracy = 0.596 (+/-0.042) for {'metric': 'minkowski', 'n_neighbors': 15}
```

L'accuracy du modèle **Sklearn** est proche de celle de **from Scratch**  
les deux modèles sont quasiment équivalents.