

TP : Projet Final

Le but de ce TP est d'implémenter un classificateur en s'appuyant sur plusieurs modélisations : réseaux de neurones (obligatoire) plus un modèle entre Naïve Bayes, SVM ou les arbres de décision. Vous disposez d'un corpus avec différentes conditions (plus d'infos dans les dossiers des data sets).

Modalités

- Choisir un data set et indiquer [ici](#) sur quel data set vous allez travailler :
<https://docs.google.com/spreadsheets/d/1BhjE6fblAUysoNdw74hi4zrmluM4ZHxBRyNNOuMHCLw/edit?usp=sharing>
 - o L'idée est d'avoir une répartition d'environ 50-50 des groupes sur les deux data sets.
- **Travail en binôme** avec partie individuelle :
 - o Ensemble (~5 périodes) :
 - Dataset exploration
 - Extraction des caractéristiques
 - Implémentation, validation et test du modèle *neural network*
 - Rapport/readme – partie 1 (en binôme)
 - o Individuel (~3 périodes) :
 - Chaque étudiant choisit un-deux modèles de machine learning additionnels parmi ceux que nous avons vu pendant le semestre et cherchera de l'optimiser
 - Extraction d'autres caractéristiques (si besoin)
 - Rapport/readme – partie 2 (individuel)
- Dans un document rapport/readme, pour chaque modèle :
 - o Expliquer l'approche pour l'extraction de caractéristiques
 - o Expliquer la séparation des données (train val test)
 - o Expliquer brièvement l'approche utilisée pour obtenir les résultats précédents
 - o Résumer les performances obtenues
 - o Indiquer les meilleurs hyperparamètres
- Vous devez appliquer (quand possible) les bonnes pratiques étudiées dans le semestre (pipelines de scikit, grid search, métriques d'évaluation pertinentes, etc.)

L'idée est de travailler sur le projet dans les ~8-9 périodes de cours. Vous avez jusqu'à **mercredi 05/05/2019 @ 10h30** pour rendre votre travail dans *cyberlearn* (1 rendu par groupe est suffisant – l'utilisation d'un gitlab/github est possible).

Bon travail !