U.E. Datamining - Machine learning - Master 2 MAS
Projet final

Voici quelques points structurants pour votre projet.

Dans chacune des étapes pensez à

- fournir une **narrative** qui décrit/explique votre démarche et offre vos conclusions (penser à un mini rapport de stage),
- utiliser des visualisations si possible,
- fournir des résultats reproductibles (surtout dans l'étape 5, la partie essentielle de votre travail), c.à.d. rédiger votre *notebook* de façon à ce qu'un lecteur puisse obtenir les mêmes résultats en l'exécutant automatiquement (`Menu` → `Kernel` → `Restart & Run All`).

Rendez les résultats sous forme d'un seul fichier jupyter (.ipynb) qui contient l'ensemble de votre travail. On le soumettra sur l'ENT (le jeu de données n'est pas nécessaire, un lien de téléchargement précis suffira).

1. Décrire les objectifs et les données. Préciser la source des données (lien internet, ...).
2. Explorer, visualiser les données, calculer des statistiques de base (pandas).
3. Scinder le jeu de données en ensemble d'apprentissage et celui de test (validation).
4. Choisir au moins deux modèles dont au moins un modèle ensembliste ; Ici, le modèle de regression simple (linéaire / logistique) ne compte pas, mais il peut éventuellement servir, si vous voulez, en modèle supplémentaire de comparaison (*base benchmark*).

   Trouver les meilleurs hyper-paramètres à l'aide de cross-validation. Pensez à bien choisir votre cross-validateur (possiblement, en utiliser deux).

   Pour tester un éventail de paramètres, on pourra faire un programme "à la main" ou utiliser `gridSearch`.
5. Conclure en appliquant le meilleur modèle à l'ensemble de test. Discuter les résultats (matrice de confusion si classification, comparer l'erreurs dans l'apprentissage et test...)