

Intelligence Artificielle

Projet Transverse

Jean-Claude Houbart

houbart.jc+epsi@gmail.com

Programme 15/04

- Constituer des équipes de 4. Pas plus de deux profils réseaux par équipe.
- M'envoyer sur houbart.jc+epsi@gmail.com la composition de votre équipe avec le profil de chacun.
- Créer votre compte sur <https://www.drivendata.org/>
 - Nom du compte initiale (x4) + _EPSI
 - Exemple : LTJCHDRPT_EPSI
- Vous inscrire sur la compétition « Pump it Up: Data Mining the Water Table »
- Récupérer les données. Ces données sont aussi utilisées sur le projet Big Data.
- NB : la description est dans [Problem Description](#)
- Analyser les données. Vous pouvez utiliser le projet Big Data.

Programme 17/04 ou 18/04

- Si nécessaire, réinstaller l'environnement Python / Jupyter Lab.
- Identifier les données incomplètes (Attention les données manquantes peuvent avoir des valeurs 0).
- Analyser l'impact des différentes variables sur le label `status_group`. Exemples dans corrigé TP ANN.
- Choisir les colonnes et compléter les données manquantes pour pouvoir les utiliser.
- Réfléchir aux modèles et librairies à utiliser pour chercher à prédire l'état du point d'eau.

Programme 16/05 et 24/05

- Faire un compte-rendu de 30 minutes à 1 heure.
- Développer les modèles de votre choix.
- Mesurer et minimiser l'overfitting.
- Faire tourner sur les données Test et publier sur drivendata.org.
- Optimiser son score.

Programme 17-18/06 et 20/06

- Faire un compte-rendu de 30 minutes à 1 heure.
- Soit continuer à développer des modèles.
- Soit constituer des modèles d'ensemble en échangeant avec les autres équipes.
- Faire tourner sur les données Test et Publier sur drivendata.org.
- Préparer un document décrivant la méthode, le publier sur DrivenData.org si le score est suffisant.

Programme 26 au 28/06

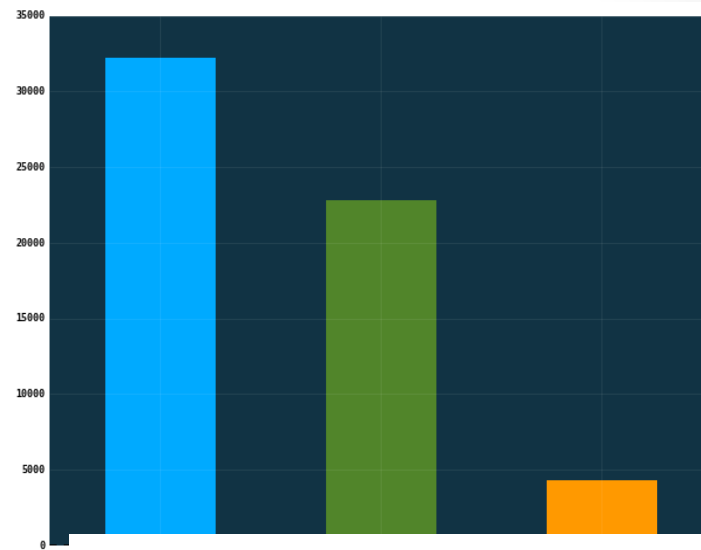
- Présentation Orale de votre projet.

Notation :

- Nettoyage et analyse des données : 5 points.
- Mise en œuvre de modèles : 8 points.
- Score sur DrivenData: 4 points.
- Présentation finale : 3 points.

Les données

- 59 400 points d'eau.
- 40 colonnes
- 3 états :
 - functional
 - non functional
 - functional need repair



Functional Not fonctionnal Need Repair

