

Collège de Bois de Boulogne

420-BD8-BB

Approche supervisée

Projet (pour le 08 Mai 2023)

Dans le cas de l'apprentissage supervisé, on développe un projet complet en se basant sur trois algorithmes de type d'apprentissage supervisé. Commencez par choisir vos données.

Étapes :

1. On s'intéresse à développer un modèle de prédiction de classes. Identifier d'abord les classes disponibles dans le dataset que vous avez choisi.
2. Procéder à une exploration du dataset et identifier des valeurs manquantes, inconnues, etc. Procéder au remplacement de ces valeurs s'il y'a lieu.
3. Développer la matrice de corrélation. Interpréter vos résultats.
4. Procéder à la séparation train/test (80-20) de votre dataset original (seed 44).
5. Utilisation et comparaison de différents modèles
 - Modèle Naive Bayes/Logistic Regression
 - Decision Tree
 - KNN
6. Calculer l'exactitude (Accuracy) sur la base des valeurs test.
7. Sauvegarde du meilleur modèle pour utilisation ultérieure
8. Utilisation du modèle sauvegardé pour la prédiction

Pour chaque étape, montrer votre démarche à l'aide des screenshots de vos commandes Python, ainsi que les résultats obtenus.

À remettre avant le 08 Mai 2023 :

- Présentation ppt du projet
- Rapport sur le projet incluant une description de détaillée de l'approche utilisée pour la réalisation de chaque étape ainsi que les résultats obtenus.

Ce travail peut se faire par des groupes de 2 à 3 personnes qui présenteront leur travail le 08 Mai en classe (20 min par présentation).

Références :

<https://moncoachdata.com/blog/nettoyage-de-donnees-python/>

<http://www.neural.cz/dataset-exploration-boston-house-pricing.html>

<http://www.analyticsvidhya.com/blog/2016/01/12-pandas-techniques-python-data-manipulation/>