

Enseignant(s)

BALASOUPRAMANIANE Amrta Devy

Email(s)

abalasoupramaniane1@myges.fr

Data Mining Python - Classe 2 - 2024

1 Matières, formations et groupes

Matière liée au projet :

Formations : -

Nombre d'étudiant
par groupe : **2 à 3**

Règles de constitution des groupes: **Libre**

Charge de travail
estimée par étudiant : **6,00 h**

2 Sujet(s) du projet

Type de sujet : **Imposé**

Projet Data Mining

Ce projet se fait par groupe de 3 (2 si pas le choix).

Deux choix de projets s'offre à vous :

- L'implémentation d'un algorithme de classification
- L'implémentation d'un algorithme de régression

Ces deux techniques de Data mining vous seront présenté en cours et sera pratiqué via TPs fait en cours.

Ce projet se constitue de :

- d'une étape de suivi avec le Choix du projet , Validation du Dataset , et commencement du nettoyage du Dataset

et

- Rendu final sous format format zip avec en contenu :

--> le Dataset

--> votre notebook

Votre présentation durera environ 15 min durant lequel vous seriez amener à faire :

- une présentation du Dataset : donc expliquer le types de données que contient votre dataset et son contexte

- Explication de votre choix de type de projets.

- une Exploration avec Visualisation : Appliquer quelques commandes statistique avec la librairie pandas, avec un graphique à l'aide des différentes librairie python de visu (ggplot, matplotlib , plotly)

- Implémentation de la Technique de Data mining avec sci-kit Learn : Nettoyage et application du modèle en expliquant votre choix et sa performance

La note finale portera sur la qualité de votre présentation ainsi que le résultat et la qualité du projet. Une démonstration sera attendu.

Pour le choix du dataset , vous êtes libre sous conditions de validation. Des datasets seront fourni au deuxième cours pour les groupes n'en ayant pas trouvé à ce moment là.

3 Détails du projet

Objectif du projet (à la fin du projet les étudiants sauront réaliser un...)

Savoir utiliser Python pour

- manipuler un jeu de données ,
- l'explorer
- le nettoyer,
- extraire de l'information en fonction du besoin en utilisation un modèle de Machine Learning

Descriptif détaillé

Savoir implémenter un modèle de classification ou de régression en utilisant les techniques de Data mining vues en cours.

Ouvrages de référence (livres, articles, revues, sites web...)

-

Outils informatiques à installer

Google Collabs (en ligne)

4 Livrables et étapes de suivi

1	Etape intermédiaire	Choisir son Dataset Faire le choix entre l'implémentation d'un algorithme de classification ou d'un algorithme de régression. Commencement le nettoyage du Dataset	mercredi 29/11/2023 23h59
2	Rendu final	Rendu Final: Format zip	samedi 06/01/2024 23h59

5 Soutenance

Durée de présentation par groupe :	15 min	Audience : A huis clos
Type de présentation :	Présentation / PowerPoint - Démonstration	
Précisions :		