



Big Data et Data Mining



Niveau d'étude
BAC +5



Composante
Polytech Dijon

Présentation

Description

Ce cours couvre les thèmes classiques et émergents de traitement et de fouille de données massives en intégrant trois parties principales : les questions conceptuelles et historiques, les principes d'utilisation des algorithmes et leur développement, ainsi que le traitement éthique de données y compris la sécurité et la vie privée. Le cours comprend les éléments suivants : l'histoire et philosophie du Big Data et Data Mining; l'extrait, transformation, chargement et nettoyage de données massives ; stockage et modélisation de données massives ; la complexité des données et rappel sur la complexité des algorithmes ; langages et outils Big Data et Data Mining; méthodes classiques d'analyse de données et les méthodes adaptées à de gros volumes de données, e.g. Partitionnement de données (K-means), réduction de dimensionnalité (PCA), map reduce, streaming ; visualisation des données massives ; traitement éthique de données, sécurité et la vie privée, anonymisation des données.

Objectifs

- Comprendre les principes de traitement et de fouille de données massives
- Utiliser des algorithmes du Big Data et Data Mining, améliorer leur compréhension
- Appliquer les méthodes d'analyse de la complexité des algorithmes et des données massives
- Visualiser et présenter les résultats
- Savoir lire, comprendre et rédiger le code, la documentation et les rapports techniques, présenter les résultats
- Comprendre l'importance d'analyse des données massives, le traitement éthique, le respect de la vie privée, RGPD, connaître les risques associés

Heures d'enseignement

CM	Cours Magistral	14h
TD	Travaux Dirigés	10,5h
TP	Travaux Pratiques	14h

Pré-requis obligatoires



- Algèbre linéaire
- Langage de programmation, e.g. Python

Modalités de contrôle des connaissances

Évaluation initiale / Session principale

Type d'évaluation	Nature de l'évaluation	Durée (en minutes)	Nombre d'épreuves	Coefficient de l'évaluation	Note éliminatoire de l'évaluation	Remarques
CC (contrôle continu)	Ecrit sur table			1.5		
CC (contrôle continu)	Production technique			1.5		

Seconde chance / Session de rattrapage

Type d'évaluation	Nature de l'évaluation	Durée (en minutes)	Nombre d'épreuves	Coefficient de l'évaluation	Note éliminatoire de l'évaluation	Remarques
CC (contrôle continu) 2nde chance	Ecrit sur table			1.5		
CC (contrôle continu) 2nde chance	Production technique			1.5		