

CERTIFICAT

SCIENCES DES DONNÉES ET BIG DATA OUTILS ET INTRODUCTION



Université
Fédérale
Toulouse
Midi-Pyrénées

DATE	PROGRAMME*	FORMAT	HORAIRES	LIEU
jeudi 8 septembre 2022	Réunion de rentrée	Présentiel/ en ligne	14h-16h	-
MODULE DE SENSIBILISATION				
lundi 12 septembre 2022	Séance d'introduction à la partie e-learning	Présentiel	18h30-19h30	INSA + en ligne (pour étudiants hors Toulouse)
Du 12 au 28 septembre 2022	M1.1 Introduction à R / <i>Exercice de manipulation de variables aléatoires et leurs distributions statistiques</i> M1.2 Elements d'analyse statistique de données : statistique élémentaire, descriptive unidimensionnelle et bidimensionnelle, estimation et tests statistiques/ <i>Autoévaluation avec jeu de données du pic d'ozone de la ville de Rennes</i> M1.3 Modélisation statistique : modèle gaussien-régression linéaire multiple, modèle binomial-régression logistique) / <i>Autoévaluation avec prédiction du pic d'ozone de la ville de Rennes</i>	E-learning	-	-
jeudi 29 septembre 2022	Séance de discussion questions/réponses sur les contenus de la première phase d'e-learning	Présentiel	14h30-16h30	INSA + en ligne (pour étudiants hors Toulouse)
Du 30 septembre au 30 octobre 2022	M1.4 Introduction à Python M1.5 Exploration multidimensionnelle : data mining, ACP-AFD, clustering avec k-means / <i>Tutoriel en autonomie : intro ACP-AFD, adaptation statistique et prévision de la concentration en ozone (exploration); segmentation d'image et cartographie de la surface de Mars</i> M1.6 Principes de l'apprentissage statistique / <i>Tutoriel en autonomie : classification supervisée de données synthétiques dans R² et adaptation statistique et prévision de la concentration en ozone (apprentissage et prévision)</i>	E-learning	-	-
mercredi 2 novembre 2022	TP sur l'exploration multidimensionnelle / tutoriel : reconnaissance d'activité à partir des enregistrements d'un smartphone (exploration)	Présentiel	9h30-12h30	INSA + en ligne (pour étudiants hors Toulouse)
	TP sur l'apprentissage statistique / tutoriel : reconnaissance d'activité à partir des enregistrements d'un smartphone (apprentissage et reconnaissance)	Présentiel	14h-17h	INSA + en ligne (pour étudiants hors Toulouse)
jeudi 3 novembre 2022	M1.7 Séance sur les algorithmes classique d'optimisation (introduction sur l'optimisation, algorithmes classiques avec gradient et spark)	Présentiel	9h-11h	INP-ENSEEIH + en ligne (pour étudiants hors Toulouse)
	TP programmer 1 algorithme	Présentiel	12h-14h	INP-ENSEEIH + en ligne (pour étudiants hors Toulouse)
vendredi 4 novembre 2022	M1.8 Séance sur les techniques de virtualisation et containerisation; les plateformes cloud	Présentiel	9h-11h	INP-ENSEEIH + en ligne (pour étudiants hors Toulouse)
	TP Déploiement dans une infrastructure virtualisée	Présentiel	12h-14h	INP-ENSEEIH + en ligne (pour étudiants hors Toulouse)
Du 10 novembre 2022 au 31 janvier 2023	M1.9 Projet de fin de module par binômes "Développer un algorithme sur une infrastructure"	Projet		
Du 10 novembre 2022 au 31 janvier 2023	Séances de suivi du projet (en soirée, 3-4 sessions)	Distanciel		En ligne
MODULE IMMERSION				
mardi 8 novembre 2022	M2.1 Séance sur les infrastructures logicielles pour le big data: MapReduce, démystification Hadoop et Spark	Présentiel	18h-20h	INP-ENSEEIH + en ligne (pour étudiants hors Toulouse)
jeudi 10 novembre 2022	TP avec Spark	Présentiel	13h - 18h	INP-ENSEEIH + en ligne (pour étudiants hors Toulouse)
jeudi 24 novembre 2022	M2.2 Séance d'introduction aux méthodes d'apprentissage : arbre binaire de décision, réseaux de neurones, support vector machines, agrégation d'arbres	Présentiel	13h-17h	ISAE-SUPAERO + en ligne (pour étudiants hors Toulouse)
jeudi 1er décembre 2022	TP Arbre binaire de décision	Présentiel	14h-17h	ISAE-SUPAERO + en ligne (pour étudiants hors Toulouse)
jeudi 8 décembre 2022	Séance introductive à la partie e-learning sur les méthodes d'apprentissage	Distanciel	14h-15h30	En ligne
Du 8 décembre 2022 au 4 janvier 2023	M2.3 Les méthodes d'apprentissage : arbre binaire de décision, réseaux de neurones, support vector machines, agrégation d'arbres	E-learning	-	
jeudi 5 janvier 2023	TP spam	Présentiel	14h-17h	ISAE-SUPAERO + en ligne (pour étudiants hors Toulouse)
jeudi 2 février 2023	M2.4 Séance sur les algorithmes stochastiques plus sophistiqués: optimisation parcimonieuse, factorisation non négative de matrice	Présentiel	14h-16h	ISAE-SUPAERO + en ligne (pour étudiants hors Toulouse)
	TP programmation d'un algorithme de chaque type	Présentiel	16h-18h	ISAE-SUPAERO + en ligne (pour étudiants hors Toulouse)
Du 3 février au 31 mars 2023	M2.5 Projet de fin de module par équipes de 4 : challenge Kaggle	Projet	-	
Du 3 février au 31 mars 2023	Séances de suivi du projet (en soirée, 3-4 sessions)	Distanciel		en ligne
CONFERENCES				
mardi 18 octobre 2022			18h30-20h30	à définir
mardi 22 novembre 2022			18h30-20h30	à définir
mardi 17 janvier 2023			18h30-20h30	à définir
mardi 7 février 2023			18h30-20h30	à définir
mardi 21 mars 2023			18h30-20h30	à définir

*ce programme est susceptible d'être modifié