

PLANNING DU CERTIFICAT DATA SCIENTIST (BD 16)

MODULES	PERIODE	DUREE
Module 1	Du 23 au 25 mars 2020	3 jours
1. Présentation du métier de data science. Rappels R	23 mars 2020	
2. Architecture de stockage des données. SQL. NoSQL	24 mars 2020	
3. Distribution stockage et calcul. Hadoop. Hive. Cloud	25 mars 2020	
Module 2	Du 27 au 29 avril 2020	3 jours
1. R et connexion aux bases de données	27 avril 2020	
2. Machine learning 1 : Régressions. Régressions logistiques	28 avril 2020	
3. Machine learning 1 : Régressions sous contraintes	29 avril 2020	
Module 3	Du 25 au 27 mai 2020	3 jours
1. Clustering et réduction dimensions	25 mai 2020	
2. Visualisation. Shiny	26 mai 2020	
3. Projet tutoré (matin). Sécurité informatique (après-midi)	27 mai 2020	
Module 4	Du 29 juin au 1^{er} juillet 2020	3 jours
1. Machine learning 2 : Boosting. Svm	29 juin 2020	
2. Machine learning 2 : Optimisation. Réseau de neurones. Deep learning	30 juin 2020	
3. Machine learning 2 : Réseau de neurones avec Word2Vec. Projet tutoré (après-midi)	1 juillet 2020	
Module 5	Du 7 au 9 septembre 2020	3 jours
1. Introduction au text mining	7 septembre 2020	
2. Passage à l'échelle	8 septembre 2020	
3. Projet tutoré (matin). Environnement juridique (après-midi)	9 septembre 2020	
Module 6	Du 12 au 14 octobre 2020	3 jours
1. Introduction au graph mining	12 octobre 2020	
2. Pipeline de traitement Python	13 octobre 2020	
3. Projet tutoré (matin). Data camp (après-midi)	14 octobre 2020	

L'examen aura lieu le **16 novembre 2020** en présentiel à l'Ensaie-Ensaï Formation Continue (Cepe).

L'ordre et le détail des modules peuvent être amenés à évoluer légèrement.