

Diplôme d'études supérieures spécialisées en science des données - 1455

CRÉDITS :

30 crédits, Deuxième cycle

DIPLÔME :

Diplôme d'études supérieures spécialisées en science des données

OBJECTIFS :

Objectif général

• Développer des compétences en science des données.

Objectifs spécifiques

• Développer une compréhension approfondie des concepts clés de la science des données :

- Apprentissage automatique et profond, visualisation de données, analyses mathématiques et statistiques;
- Appliquez des techniques de modélisation et d'analyse de données pour la résolution de problèmes concrets;
- Acquérir une expérience pratique en langages de programmation et outils en science des données;
- Communiquer les résultats des analyses à l'aide de techniques et d'outils de visualisation;
- Comprendre les enjeux de la gouvernance de données dans les organisations.

INFORMATION SUR L'ADMISSION :

Lieu d'enseignement	Régime	Trimestres d'admission			Étudiants étrangers		
		Aut.	Hiv.	Été	Aut.	Hiv.	Été
Campus de Rimouski	TC	✓			✓		
	TP	✓					
Campus de Lévis	TC	✓			✓		
	TP	✓					

TC : Temps complet

TP : Temps partiel

CONDITIONS D'ADMISSION :

Base études universitaires

La candidate ou le candidat doit être titulaire d'un baccalauréat, ou l'équivalent, en informatique, en informatique de gestion ou en génie informatique, obtenu avec une moyenne cumulative d'au moins 3 sur 4,3 ou l'équivalent.

Base expérience

Posséder les connaissances requises, une formation appropriée et une expérience jugée pertinente.

Les méthodes et les critères de sélection consistent à l'évaluation du dossier scolaire et des lettres de recommandation. Les candidates et les candidats qui présentent une demande sur la base de l'expérience pertinente seront convoqués à une entrevue. Avant d'être admis au programme, la candidate ou le candidat peut être appelé à parfaire sa formation par une propédeutique ou par des cours d'appoint.

PLAN DE FORMATION :

Bloc 1

Quatre (4) cours obligatoires :

INF70022	Introduction à la science des données (3 cr.)
INF70122	Introduction à l'intelligence artificielle (3 cr.)
INF70222	Programmation pour la science des données (3 cr.)
MAT70322	Mathématiques et statistiques pour la science des données (3 cr.)

Cours optionnels

Quatre (4) cours parmi les blocs 2, 3 et 4.

Bloc 2 Cours de spécialité

Deux (2) à trois (3) cours parmi les suivants (6 à 9 crédits) :

INF71022	Analyse sémantique et traitement automatique de la langue naturelle (3 cr.)
INF71122	Optimisation avec les métaheuristiques (3 cr.)
INF71222	Visualisation de données et « reporting » numérique (3 cr.)
INF71322	Gestion et traitement de mégadonnées (3 cr.)

INF71422	Apprentissage automatique (3 cr.)
INF71522	Apprentissage profond I (3 cr.) (INF71422)
INF71622	Apprentissage profond II (3 cr.) (INF71522)
INF71722	Ingénierie d'apprentissage automatique pour la production (MLOps) (3 cr.) (INF71422)
INF71822	Gouvernance des données (3 cr.)
INF71922	Qualité des données (3 cr.)
INF73022	Introduction à l'intelligence d'affaires (3 cr.)
INF73122	Sujets spéciaux en sciences de données (3 cr.)

Bloc 3 Cours d'application

Un (1) ou deux (2) cours parmi les suivants (3 à 6 crédits) :

INF72022	Vision numérique (3 cr.) (INF71422)
INF72122	Analyse de séries temporelles (3 cr.) (INF71422)
INF72222	Intelligence artificielle pour les jeux vidéo (3 cr.) (INF71422)
INF72322	Intelligence artificielle pour la cybersécurité (3 cr.)
INF72422	Santé intelligente (3 cr.) (INF71422)
INF72522	Intelligence artificielle pour l'industrie 4.0 (3 cr.) (INF71422)

Bloc 4 Cours d'enrichissement

Zéro (0) ou un (1) cours d'enrichissement pourra être suivi dans un autre programme de deuxième cycle de l'UQAR, sur approbation de la direction du programme.

Bloc 5 Activité finale

Une activité (6 crédits) parmi les suivantes :

INF79022	Stage (6 cr.) (INF71422)
INF79122	Projet (6 cr.)

Programme approuvé par la Commission des études du 4 octobre 2022 (CE-784-7553).

INF70022**Introduction à la science des données**

Objectif : Introduire aux concepts de base de la science des données.

Contenu : Disciplines de la science des données. Démarche, méthodes et processus de la science des données. Défis et enjeux de la science des données. Outils, langages et bibliothèques communs utilisés en science des données. Applications de la science des données.

INF70122**Introduction à l'intelligence artificielle**

Objectif : Introduire aux concepts clés, aux principes, et aux techniques de l'intelligence artificielles.

Contenu : Techniques de représentation des connaissances et raisonnements : agents logiques (réactifs, rationnels, intelligents, autonomes), différents types de logiques (entre autres : logique des propositions, logique du premier ordre, logiques d'ordres supérieures, logique floue), raisonnement déductif et inductif, formulation et résolution de problèmes, graphes d'états, algorithmes de recherche informée et non informée, algorithmes de jeux de stratégie, apprentissage à partir d'exemples, réseaux de neurones, algorithmes de regroupement, réseaux bayésiens, apprentissage profond, apprentissage d'ensemble, apprentissage par renforcement. Méthodologie de développement, environnements et cadres de développement, langages. Approches émergentes en intelligence artificielle.

INF70222**Programmation pour la science des données**

Objectif : Couvrir les principaux concepts, outils et langages de programmation utilisés pour résoudre des problèmes en science des données.

Contenu : Langages de programmation utilisés en science de données (entre autres : Python, R, Julia). Concepts clés pour la conception de programmes pour l'exploration et l'analyse des données : structures de données, structures de contrôle, récurrence, procédures, fonctions. Apprentissage du développement de programmes pour : nettoyer, filtrer, transformer, agréger et arranger des données. Interaction par programmes avec des bases de données. Systèmes de contrôle de versions tels que Git. Environnements de développement pour la science des données (entre autres : Jupyter et JupyterLab, Google Colab). Manipulation de grands jeux de données. Évaluation des performances d'un programme.

INF71022**Analyse sémantique et traitement automatique de la langue naturelle**

Objectif : Développer une compréhension approfondie des techniques et algorithmes utilisés pour le traitement automatique d'information linguistique et le langage naturel.

Contenu : Fondamentaux du traitement de la langue naturelle (entre autres : tokenisation, stemming, lemmatisation, stop words, reconnaissance d'entités nommées, plongements de mots, similarité entre phrases, analyse de graphes, Tf-IdF, extraction de relations, extraction d'événements). Techniques d'étiquetage grammatical (entre autres : tables de correspondance, n-grams, modèles de Markov cachés). Traitement automatique de la langue naturelle (entre autres : extraction de mots clé, analyse d'opinions/sentiments, modélisation de sujets, agents conversationnels, correction orthographique, traduction automatique). Modèles d'apprentissage automatique avancés pour le traitement de la langue naturelle (entre autres : réseaux de neurones denses et récurrents, modèles transformationnels, modèles d'attention).

INF71122**Optimisation avec les métaheuristiques**

Objectif : Acquérir une vue d'ensemble de la démarche à suivre en vue de résoudre un problème d'optimisation combinatoire théorique ou pratique complexe.

Contenu : Concepts de base (entre autres : complexité, heuristique, métaheuristique, optimisation, représentation de solutions). Démarche scientifique et problèmes d'optimisation; résolution des problèmes et présentation des résultats obtenus. Méthodes de recherche de solutions (entre autres : recuit simulé, recherche avec tabous, algorithmes génétiques, l'optimisation par colonie de fourmis, essais particuliers, recherche à voisinage variable, recherche à voisinage large adaptative). Introduction à l'optimisation multi-objectifs.

INF71222**Visualisation de données et « reporting » numériques**

Objectif : Apprendre les différentes techniques de visualisation de données.

Contenu : Visualisation de données : introduction et principes. Catégories d'affichage de la visualisation de données (entre autres : tables, tableaux de bord, graphes et courbes, cartes). Visualisation multivariée, géospatiale, et interactive. Outils et bibliothèques de visualisation de données. Bibliothèques de manipulation et de visualisation de données (entre autres : NumPy, SciPy, Pandas, Matplotlib, Seaborn, plotly, ggplot avec Python et dplyr, reshape2, ggplot2, Lattice, LeafletR, Plotly, sunburstR, RGL, dygraphs avec R). Reporting : avantages, types, démarche et règles à suivre, données à intégrer, et reporting attractif.

INF71322**Gestion et traitement de mégadonnées**

Objectif : Introduire aux concepts, techniques, et outils de stockage, de recherche et d'analyse de mégadonnées.

Contenu : Bases de données (entre autres : relationnelles, objet, objet-relationnel, NoSQL). Caractéristiques et propriétés des mégadonnées et des bases de données pour mégadonnées. Paradigmes pour la distribution de calculs et d'apprentissages automatique à partir de mégadonnées (entre autres : Map-Reduce, ensembles de données distribués résilients). Pipelines de gestion et d'analyse de données diffusées en continu. Conception d'algorithmes pour la résolution de problèmes de mégadonnées. Ressources technologiques (logiciels, cadres, bibliothèques, solutions infonuagiques) pour le stockage, la gestion et l'analyse des mégadonnées (entre autres : Hive, Flume, HDFS, Kafka, Pig, HDFS, Sqoop, Spark). Solutions infonuagiques pour les mégadonnées.

INF71422**Apprentissage automatique**

Objectif : Introduire aux concepts de base et aux techniques d'apprentissage automatique.

Contenu : Concepts clés, applications et techniques de l'apprentissage automatique. Formulation des problèmes d'apprentissage automatique correspondant à différentes applications, et développement de modèles pour résoudre ces problèmes. Entrées (concepts, instances, attributs, types des données et leur préparation). Sorties et représentation des connaissances (itemsets, règles d'association, patrons fréquents, classes, groupes). Prétraitement des données, réduction de la dimensionnalité, et sélection de caractéristiques/variables, Types d'apprentissage (entre autres : supervisé, non supervisé, semi-supervisé, par renforcement). Regroupement. Arbres de décision. Forêts aléatoires. Régression (logistique, linéaire, multiple, arbres de régression). Apprentissage d'ensemble (« boosting », « bagging », « stacking »), fédéré, par transfert, profond, et incrémental. Surapprentissage et sous apprentissage. Évaluation des modèles résultats d'un processus d'apprentissage automatique.

INF71522**Apprentissage profond I**

Objectif : Se familiariser avec les concepts de base de l'apprentissage profond.

Contenu : Aperçu de l'évolution des réseaux de neurones passant du perceptron aux architectures profondes. Avantages et défis de l'apprentissage profond. Paramétrage des réseaux de neurones. Architectures de réseaux de neurones pour l'apprentissage profond (entre autres : RNN, CNN, LSTM, SOM, GAN, GRU). Ressources technologiques (cadres, bibliothèques, solutions infonuagiques) : développement et déploiement de modèles d'apprentissage basés sur des architectures profondes.

INF71622**Apprentissage profond II**

Objectif : Découvrir les concepts avancés de l'apprentissage profond.

Contenu : Interprétabilité et explicabilité des modèles d'apprentissage profond. Apprentissage profond multivues et avec les réseaux à impulsions. Utilisation d'outils et de bibliothèques de référence. Nouvelles tendances en apprentissage profond.

INF71722**Ingénierie d'apprentissage automatique pour la production (MLOps)**

Objectif : Concevoir des systèmes de production d'apprentissage automatique de bout en bout.

Contenu : Pipelines, contraintes et exigences de déploiement. Besoins en données et pipelines de données. Stratégies de modélisation. Identification des technologies de base nécessaires pour la prise en charge des MLOps efficaces. Bonnes pratiques CI/CD dans le contexte des systèmes ML. Conception, construction et maintenance des systèmes de production de modèles d'apprentissage automatique.

INF71822**Gouvernance des données**

Objectif : Comprendre les concepts de base de la gouvernance de données et son importance.

Contenu : Importance des actifs numériques et leur valorisation dans les organisations. Missions, principes, politiques, stratégies, et bénéfices de la gouvernance de données. Évaluation de la maturité. Cadre de gouvernance. Rôles et responsabilités. Processus et programmes. Bonnes pratiques. Étude de cas de gouvernance de données dans différents types d'organisations.

INF71922**Qualité des données**

Objectif : Maîtriser les processus et techniques de base pour identifier et résoudre des problèmes de qualité des données.

Contenu : Connaissance des données (entre autres : complexité, qualité, quantité, actualisation, types, formats, sources). Identification et résolution des problèmes de qualité des données. Préparation et prétraitement des données (entre autres : profilage, filtrage, nettoyage, standardisation et normalisation, transformation, ingénierie des caractéristiques, réduction de dimension). Exigences, critères, et métriques de qualité des données.

INF72022**Vision numérique**

Objectif : Maîtriser les notions de base de la vision numérique 2D et 3D par ordinateur.

Contenu : Fondements de la vision artificielle. Prétraitement des images et des vidéos. Détection et extraction de caractéristiques. Segmentation. Applications en classification, reconnaissance de formes et d'objets,

estimation et suivi des mouvements. Outils et techniques de la vision par ordinateur. Utilisation de l'apprentissage automatique classique pour la vision par ordinateur (entre autres : KNN, régression logistique, SVM, apprentissage profond). Applications de la vision par ordinateur.

INF72122

Analyse de séries temporelles

Objectif : S'initier aux concepts théoriques et approches pratiques de développement de modèles d'apprentissages automatiques appliqués aux séries temporelles.

Contenu : Fondamentaux de l'analyse des séries temporelles : données temporelles, taxonomie des problèmes de prévision, développement de modèles de prévision, transformation de problème de données temporelles en problème d'apprentissage supervisé, méthodes classiques de prédiction. Principales méthodes classiques et d'apprentissage profond pour la prédiction avec les séries temporelles univariées et multivariées (entre autres : ARIMA, SARIMA, Auto-ARIMA, ETS, VARMA, LSTM). Applications de l'analyse de séries temporelles.

INF72222

Intelligence artificielle pour les jeux vidéo

Objectif : Comprendre les principaux enjeux de l'utilisation de l'intelligence artificielle dans les jeux vidéo.

Contenu : Principes de la conception des jeux. Principales approches d'intelligence artificielle et types de jeux (entre autres : algorithmes génétiques, apprentissage par renforcement). Décisions tactiques et stratégiques. Coordination d'agents. Génération de contenu. Apprentissage des techniques.

INF72322

Intelligence artificielle pour la cybersécurité

Objectif : Utiliser l'intelligence artificielle pour résoudre des problèmes de cybersécurité.

Contenu : Besoins essentiels en cybersécurité : identification, authentification, confidentialité, autorisation, intégrité, non répudiation. Cadre de contrôle pour la cybersécurité et la gestion de risques : identification, protection, détection, réponse, rétablissement. Techniques d'apprentissage classiques et profond pour : modélisation de comportements, détection de patterns, identification d'anomalies, détection d'intrusions, prédiction de risques de brèches de sécurité, détection de nouvelles menaces.

INF72422

Santé intelligente

Objectif : S'initier aux techniques d'intelligence artificielle appliquée à la santé.

Contenu : Concepts fondamentaux : défis en santé, concepts de santé

intelligente, la santé intelligente et l'éthique. Opportunités actuelles et futures en matière de santé intelligente. Développement de solutions à des défis spécifiques (entre autres : diagnostique, détection). Conception, implantation, évaluation et intégration des modèles prédictifs. Application de techniques de l'intelligence artificielle à différents types de données : imagerie médicale 2D et 3D, données EHR, données de dispositifs portables (« wearable devices »).

INF72522

Intelligence artificielle pour l'industrie 4.0

Objectif : Comprendre les technologies clés stimulant le développement de la fabrication intelligente en industrie 4.0.

Contenu : Industrie 4.0 : contexte historique et évolution du concept. Principes et objectifs opérationnels de l'industrie 4.0. Technologies clés stimulant le développement de la fabrication intelligente en industrie 4.0 : l'internet des objets, l'infonuagique, et l'intelligence artificielle. Intégration cyberphysique (systèmes cyberphysiques et les jumeaux numériques). Utilisation de l'intelligence artificielle en industrie 4.0 pour la planification, la production et le contrôle : prédiction de la qualité et du rendement, maintenance prédictive, collaboration homme-robot, conception générative, optimisation des chaînes d'approvisionnement.

INF73022

Introduction à l'intelligence d'affaires

Objectif : Maîtriser les concepts de base de l'intelligence d'affaires.

Contenu : Principes et concepts de base en intelligence d'affaires. Méthodologie, stratégies et ressources technologiques pour le développement de solutions d'intelligence d'affaires. Intégration de l'intelligence d'affaire dans la prise de décisions au sein des organisations. Intelligence d'affaires et entrepôts de données (« data warehouses »), magasins de données (« data marts »), et lacs de données (« data lakes »). Utilité de l'intelligence d'affaires. Apprentissage automatique en intelligence d'affaires.

INF73122

Sujets spéciaux en sciences de données

Objectif : Avoir accès à divers domaines spécialisés ou nouveaux en sciences de données.

Contenu : Contenu variable en fonction des besoins des étudiantes et des étudiants, de l'expertise professorale disponible et du domaine de spécialisation couvert.

INF79022

Stage

Objectif : Se familiariser avec des problèmes concrets en science de données dans un milieu professionnel.

Contenu : Stage en science de données à temps plein dans un milieu professionnel d'une durée de 4 à 6 mois répartis sur un ou deux trimestres, équivalent à 270 heures de travail. Complément pratique à la formation des étudiantes et des étudiants. Contenu précis déterminé en fonction des milieux professionnels.

INF79122

Projet

Objectif : Intégrer les connaissances théoriques et pratiques acquises dans la réalisation d'un projet en science de données.

Contenu : Projet en science de données à temps plein à l'université d'une durée de 4 à 6 mois répartis sur un ou deux trimestres, équivalent à 270 heures de travail. Complément pratique à la formation des étudiantes et des étudiants. Résolution de problème en science de données. Analyse des différents aspects techniques et économiques du problème soumis. Évaluation des différentes solutions et justification de celle retenue. Conception et implantation de la solution choisie. Contenu précis déterminé en fonction du projet proposé.

MAT70322

Mathématiques et statistiques pour la science des données

Objectif : Comprendre les concepts mathématiques et statistiques utilisés en science des données.

Contenu : Statistiques descriptives et théorie des probabilités : introduction, variable, population et échantillonnage, processus stochastique, mesure du centre (moyenne, médiane, mode), asymétrie, écart interquartile, variance, analyse de la variance et déviation standard, estimateurs, normalisation, calcul des probabilités, théorème de Bayes, loi des grands nombres, théorème de la limite centrale. Distributions de probabilités : discrètes (entre autres : Binomiale, Géométrique, Poisson), continues (entre autres : Exponentielle, Normale). Régression et métriques : coefficient de corrélation, R-Squared, MSE, MAE, RMSE. Estimation statistique. Test d'hypothèse.