

Plan de cours

Titre du cours : Analyse exploratoire des données

Code officiel : 420-A55-BB

Programme d'études : Spécialisation technique en intelligence artificielle

Session visée par le cours : 2^e session – groupe 01302 – Été 2025

Date :

Discipline : Techniques informatiques

Préalables :	Aucun
---------------------	-------

Préparatoire à	Aucun
-----------------------	-------

Pondération		
Nombre d'heures par semaine		
Heures/contact en classe	Heures de laboratoire ou de travail dirigé	Heures de travail à la maison ou à la bibliothèque
2	2	4
Nom du ou des enseignants de ce cours		Coordonnées : téléphone, courriel, bureau
Komi Sodoké, PhD		komi.sodoke@bdeb.qc.ca

Ce plan de cours est un guide mis à votre disposition pour organiser vos activités d'études et de travail liées à ce cours. Il est donc important que vous en preniez connaissance, de façon attentive, puisqu'il vous renseigne sur les apprentissages à réaliser, les exigences du cours et les éléments essentiels de son déroulement. **Après que l'enseignant vous l'a remis et expliqué, le contenu de ce plan de cours ne peut pas être modifié sans vous avoir consulté.**

Place du cours dans le programme d'études

Ce cours est dispensé durant la 1^{ère} session du programme et permet l'atteinte de la compétence « Effectuer l'analyse exploratoire des données ». Il permet aux étudiants d'atteindre les compétences nécessaires pour l'exploration d'un jeu de donnée en préparation à l'application d'une approche d'apprentissage automatique. Il leur permet ainsi de bien comprendre le jeu de données en appliquant différentes mesures et métriques statistiques pour l'exploration des données et d'élaborer des graphiques appropriés selon les différentes catégories des données utilisées.

Ainsi l'étudiant peut décrire plus facilement les données, après avoir identifié les différents liens qui peuvent exister entre elles et ainsi sélectionner les variables appropriées pour les algorithmes d'apprentissage automatique.

Ce cours sera divisé en trois parties, la première vise à introduire l'étudiant aux concepts de l'analyse exploratoire des données ainsi qu'aux outils et langage(s) correspondants. La seconde décrit en détail le processus d'exploration de données ainsi que les différentes mesures et métriques selon les différentes catégories de données. La troisième partie sera consacrée à l'application des différentes approches pour l'analyse exploratoire des données.

Tout cours vise l'atteinte d'une ou de plusieurs compétences prescrites par le Ministère. Une compétence fait appel à des connaissances que vous possédez déjà et vous amène à acquérir de nouveaux savoirs et à développer de nouvelles habiletés. L'atteinte de la compétence est importante pour votre développement professionnel ou pour répondre à des exigences universitaires.

Compétence(s) à atteindre dans ce cours	Éléments de compétence (principales étapes de réalisation)
Effectuer l'analyse exploratoire des données	<ol style="list-style-type: none">1 Identifier les objectifs de l'analyse exploratoire des données.2 Définir les concepts de l'analyse exploratoire des données.3 Décrire les étapes du processus de l'analyse exploratoire des données.4 Inspecter le(s) jeu(x) de données à utiliser dans le contexte de l'IA5 Évaluer la qualité des données6 Identifier les différentes catégories de données7 Reconnaître les différentes mesures et métriques pour l'exploration de données selon les différentes catégories de données utilisées8 Identifier les types de représentations graphiques adéquates aux catégories de données utilisées9 Appliquer différentes approches pour l'analyse exploratoire des données.10 Interpréter les résultats de l'analyse exploratoire des données

Planification du cours

Voici les différentes activités qui vous amèneront à atteindre la ou les compétences visées par le cours. Comme vous le verrez ci-dessous, le cours est découpé en objectifs terminaux qui traduisent le résultat attendu au terme d'une séquence d'apprentissage. Pour chacun, des objectifs spécifiques viennent spécifier ce à quoi vous devez parvenir pour atteindre l'objectif terminal. Le contenu détaillé précise les sujets amenés dans ce cours et le calendrier indique le moment où ce contenu sera traité.

Objectif terminal :	Expliquer les objectifs et les étapes de l'analyse exploratoire des données.	
Objectifs spécifiques	Contenu détaillé	Calendrier
Rappels des définitions :	<ul style="list-style-type: none"> • Approches d'apprentissage machine (IA) • Analyse exploratoire des données • Approches statistiques 	Semaine 1
Introduction à l'analyse exploratoire :	<ul style="list-style-type: none"> • Définition • Objectifs de l'analyse dans le contexte de l'apprentissage machine • Importance de la qualité des données dans le développement d'une application IA (pertinence, intelligibilité, exactitude, accessibilité, ...) 	Semaine 2
Processus d'analyse exploratoire :	<ul style="list-style-type: none"> • Définition de la notion de la caractéristique de la donnée • Analyse des besoins et sélection des données appropriées : population, échantillon, estimations, ... • Exploration des caractéristiques des données (unidimensionnelle, multidimensionnelle, séries temporelles, probabilités, etc.) • Analyse avancée des données (Composantes Principales, Correspondances simple et Multiple, ...) • Détection des valeurs atypiques (aberrante, manquantes, ...) • Interprétation des résultats • Documentation 	Semaine 3

Objectif terminal :	Expliquer les objectifs et les étapes de l'analyse exploratoire des données.	
Objectifs spécifiques	Contenu détaillé	Calendrier
Installation et configuration des logiciels dédiés à l'analyse exploratoire :	<ul style="list-style-type: none"> • Présentation des outils d'analyse exploratoire • Analyse des besoins des environnements d'analyse exploratoire dans le contexte d'une solution d'IA • Comparaison des écosystèmes de développement : performance, plateforme, langages, licences, ... • Critères de choix selon le contexte de l'étude. • Installation et paramétrage de l'environnement 	Semaine 3/4/5
Identifier les types de représentations	<ul style="list-style-type: none"> • Identifier les types de graphique • Identifier les types de données dans vos graphiques • Identifier les types de représentations graphique selon vos types de données • 	Semaine 3/4/5

Objectif terminal :	Définir les concepts de l'analyse exploratoire des données	
Objectifs spécifiques	Contenu détaillé	Calendrier
Définir les concepts de l'analyse exploratoire des données	<p>Domaine des statistiques :</p> <ul style="list-style-type: none"> • Histoire, définition, objectifs des statistiques, données statistiques. • Éléments statistiques, terminologie et notation standard • Variables, données statistiques, tableaux, effectifs, ... • Probabilités et variables aléatoires • Algèbre linéaire de base • Contexte et cas d'utilisation <p>Statistique descriptive et inférentielle</p> <ul style="list-style-type: none"> • Définition • Caractéristiques des deux approches • Notion de base : distribution, lois usuelles, probabilités, convergences, estimation, etc. • Notion de Tests d'hypothèse • Différence entre les deux approches 	Semaine 6

Objectif terminal :	Expliquer les objectifs et les étapes de l'analyse exploratoire des données.	
Objectifs spécifiques	Contenu détaillé	Calendrier
Langage de programmation :	Éléments de syntaxe du langage R <ul style="list-style-type: none"> • Structure de contrôle et boucles, fonctions, ... • Modules d'analyse exploratoire (fonctions statistique, lecture des différents formats de données, ...) • Fonctions graphiques 	Semaine 6/7
Objectif terminal :	Description de(s) jeu(x) de données selon les différentes catégories des variables	
Objectifs spécifiques	Contenu détaillé	Calendrier
Description de(s) jeu(x) de données selon les différentes catégories des variables	Principaux facteurs de la qualité (pertinence, validité, intelligibilité, exactitude, accessibilité, disponibilité...) Critères de la qualité des données : <ul style="list-style-type: none"> • Exhaustivité, fiabilité • Intégrité et Complétude • Validité des résultats d'analyse • Évaluation des hypothèses • Exemple de cas d'étude avec des données de qualité et données de mauvaise de qualité Mesures de la qualité des données <ul style="list-style-type: none"> • Valeurs manquantes, erronées ou atypiques, biais expérimentaux, modalités trop rares, doublons, ... • Échantillonnage (empirique, aléatoire, ...) • Estimation et tests statistiques • Tests d'hypothèse, de conformité • Etc. Étude de cas <ul style="list-style-type: none"> • Fonctions du langage de programmation • Implémentation des mesures de détection de la qualité • Interprétation des résultats 	Semaine 8

Objectif terminal :	Reconnaître les différentes mesures et métriques pour l'exploration de données selon les différentes catégories de données utilisées	
Objectifs spécifiques	Contenu détaillé	Calendrier
Mesures et métriques usuelles :	<ul style="list-style-type: none"> • Notion des différentes caractéristiques des variables (Tendance centrale, Dispersion, Forme, Convergences des suites de variables aléatoires, Cooccurrences, valeur propres, etc.) • Algèbre linéaires (vecteur, matrice, distance, déterminants, valeurs propres, diagonalisation, dépendance linéaire, ...) • Position (mode, médiane, moyenne, moyenne harmonique, moyenne pondérée, ...) • Dispersion (étendue, écart-type, variance, fréquences, ...) • Forme (Coefficient d'asymétrie de Pearson/Fisher) • Concentration (indice de Gini, ...) • Liaison (corrélations, covariances, cooccurrence, ...) • Probabilité (conditionnelles, marginales) • Décomposition en valeurs singulières • Etc. 	Semaine 9
Sélection des mesures et métriques pour les variables	<ul style="list-style-type: none"> • Qualitative nominale et ordinale (position, fréquences et tableau statistique, distribution, ...) • Quantitative discrète et continue (tableau statistique, fonction de répartition, ...) • Aléatoire (loi de probabilité, moments et quantiles, ...) • Chronologique (tendance, cycle, phénomène périodique, ...) • Cas des variables au format particulier (segmentation, pondération, ...) • Etc. • Fonctions et librairie implémentant des mesures et métriques statistiques pour l'analyse exploratoire des données du langage de programmation choisi • Paramètres des fonctions implémentant les mesures et métriques statistiques pour l'analyse exploratoire des données 	Semaine 10

Objectif terminal :	Appliquer différentes approches pour l'analyse exploratoire des données.	
Objectifs spécifiques	Contenu détaillé	Calendrier
Appliquer différentes approches pour l'analyse exploratoire des données.	<p>Caractéristiques des approches de l'analyse exploratoire à :</p> <ul style="list-style-type: none"> • Univariée • Bivariée et Multivariée <p>Mesures et métriques statistiques pour les différentes approches de l'analyse exploratoire :</p> <ul style="list-style-type: none"> • Univariée (Position, Dispersion, Forme, Concentration,...) • Bivariée et Multivariée (Covariance, corrélation, distribution, tableau de contingence, indice de liaison, ...) <p>Mises-en oeuvre d'une analyse exploratoire sur un jeu de données :</p> <ul style="list-style-type: none"> • Configuration des fonctions et librairies graphiques du langage de programmation choisi • Exploration Univariée • Exploration Bivariée et Multivariée <p>Rédaction d'un rapport d'analyse exploratoire des données incluant les graphiques et les résultats des mesures statistiques</p>	Semaine 11

Objectif terminal :	Synthétiser les résultats de l'analyse exploratoire des données	
Objectifs spécifiques	Contenu détaillé	Calendrier
<p>Identification des dispositifs de stockage</p> <p>Utilisation correcte des dispositifs de stockage.</p>	<p>Éléments d'interprétation :</p> <ul style="list-style-type: none"> • Qualité (données manquante, aberrante, bias...) • Observation des tendances et comportement atypique • Liaison des variables et modalités de dispersion • Validation des tests d'hypothèses statistiques • Etc. <p>Conclusions de l'analyse exploratoire appliquée au jeu de données selon l'étude de cas :</p> <ul style="list-style-type: none"> • Argumentation des résultats d'interprétation • Explication des relations possibles entre les variables (incohérences, liaisons linéaires/non linéaire, • Identification les variables candidates dans le cadre d'un apprentissage machine • Documentation des éléments d'interprétation <p>Préparation à l'étapes suivantes de l'analyse exploratoire dans le contexte de la mise en place d'une solution en IA :</p> <ul style="list-style-type: none"> • Recommandation pour amélioration de la qualité des données • Identification des variables assujetties à une transformation/nettoyage • Etc. 	Semaine 12

Cette planification demeure une projection du déroulement du cours. Celle-ci peut subir des changements, **avec préavis**.

Méthodes d'enseignement et d'apprentissage

Voici les différentes méthodes d'enseignement et d'apprentissage que l'enseignant utilisera pour vous amener à atteindre les objectifs terminaux visés par ce cours.

Démonstrations de programmes, laboratoires pratiques de programmation et exposés magistraux.

Les concepts sont expliqués en classe et programmés systématiquement en python grâce à des exemples simples et clairs.

Modalités d'évaluation

Évaluation formative

L'évaluation formative consiste à des activités dont le but est de vous aider, pendant l'apprentissage, à distinguer ce qui est acquis de ce qui ne l'est pas, de découvrir où et en quoi vous éprouvez des difficultés d'apprentissage et à proposer les correctifs appropriés.

Démonstrations de programmes, laboratoires.

Évaluation sommative

L'évaluation sommative consiste à mesurer l'apprentissage portant uniquement sur l'atteinte des objectifs terminaux du cours, pour vous donner une note au bulletin. Le minimum requis pour réussir le cours est de 60 % et la réussite du cours est obligatoire pour obtenir le diplôme.

Le jugement de l'enseignant doit se faire par des activités évaluées à partir de critères connus par vous. Ces activités ont une pondération substantielle et leur réussite est déterminante sur le cheminement futur de vos apprentissages. La décision finale demeure le jugement de l'enseignant ; c'est sa responsabilité de déterminer si vous avez atteint le seuil minimal de réussite, en tenant compte de différents facteurs lui permettant de prendre une décision.

Vous trouverez, ci-dessous, la forme des différentes activités d'évaluation ainsi que leurs critères qui permettront de juger du degré de maîtrise de la compétence. Pour chacun, la pondération et le moment de passation sont indiqués.

Forme	Critères d'évaluation	Pondération	Échéancier
Travail noté	<ul style="list-style-type: none"> • Inspecter les jeux de données à utiliser dans le contexte de l'AED • Évaluer la qualité des données • Identifier les types de représentations graphiques adéquates aux catégories de données utilisées 	30%	Fin du cours
Intra	<ul style="list-style-type: none"> • Inspecter les jeux de données à utiliser • Réaliser une AED via des méthodes factorielles 	30%	Suite à la séance 8
Examen final	<ul style="list-style-type: none"> • Inspecter les jeux de données à utiliser • Réaliser une AED via des regroupements et des règles d'associations 	40%	Suite à la séance 12

Objectifs relatifs à la langue française

La Politique de valorisation de la langue française précise des modalités pour vous assurer la maîtrise du français. Pour améliorer la qualité de la langue, Le professeur s'assurera que les communications écrites respectent les règles de la langue française. Il contrôlera l'application de ces règles lors des différentes évaluations. Dans le cas des travaux de programmation, les identificateurs devront être des termes en français, les commentaires ainsi que les textes apparaissant à l'écran lors de l'exécution des programmes devront être dans un français convenable. Jusqu'à 5% de la note pourrait être perdu à ce titre.

Règles particulières

Règles particulières

Ces règles sont des exigences particulières que vous devez respecter pour assurer votre réussite et l'atteinte de la ou des compétences visées par ce cours.

Présence aux cours et aux activités d'apprentissage	La présence au cours et aux activités d'apprentissage est obligatoire. Vous ne pouvez pas décider selon votre bon vouloir d'assister ou non aux cours. Votre participation active aux différentes activités d'apprentissage réalisées en classe est un des facteurs importants permettant l'acquisition progressive de la compétence visée et d'augmenter vos chances de réussite : les exposés du professeur, les discussions en groupe, les explications et les exemples donnés oralement, les échanges sur les productions des autres élèves, les évaluations formatives, etc., font partie intégrante de la démarche d'apprentissage prévue pour être menée en classe et non selon un modèle de formation à distance. (PIEA, 2019 p. 6)
Absence à une épreuve sommative	Dans le cas d'une absence justifiée par une raison valable (par exemple : maladie, accident ou décès dans la famille proche), l'étudiant communique avec son enseignant dans les délais prévus dans le plan de cours pour fixer les modalités et le moment de la passation d'une épreuve équivalente. Une absence non justifiée à cette reprise entraîne la note zéro. Dans les cas d'absences prévues et justifiées, l'enseignant et l'étudiant conviennent d'avance des accommodements possibles. L'étudiant qui, sans raison valable, ne se présente pas à une épreuve se voit attribuer la note zéro pour celle-ci. (PIEA, 2019 p. 7)
Respect des échéances	Chaque journée de retard dans la remise d'un travail entraîne une diminution de 5 % de la note attribuée. Tout travail non remis à l'enseignant dans un délai de trois jours ouvrables après l'échéance n'est pas considéré et la note zéro est attribuée. L'enseignant refuse le travail d'un étudiant si les autres membres de la classe possèdent déjà les résultats de l'évaluation de ce même travail : la note zéro est alors octroyée. Quand une raison valable justifie le retard (par exemple : maladie, accident ou décès dans la famille proche), l'étudiant s'entend avec l'enseignant sur les modalités de remise de son travail. (PIEA, 2019, p. 8)
Précisions relatives à la fraude et au plagiat	Toute fraude (incluant le plagiat), toute tentative de fraude ou toute collaboration à une fraude avant ou pendant une épreuve est sanctionnée selon les modalités rendues publiques par le Collège. Selon la nature de la fraude ou du plagiat, les sanctions suivantes peuvent s'appliquer : <ul style="list-style-type: none"> • la note zéro pour la partie de l'épreuve ou du travail • la note zéro pour l'épreuve ou le travail • l'échec au cours, si la valeur pondérée de l'épreuve ou du travail le justifie et, s'il y a récurrence, <ul style="list-style-type: none"> • la suspension du Collège pour une session • l'expulsion du Collège (PIEA, 2019, p. 9)

Règles particulières

Médiagraphie

Ces documents de base sont des références qui peuvent vous être utiles dans vos activités d'apprentissage.

Titre :

Exploratory Data Analysis Using R Paperback – Sep 4 2018 by Ronald K. Pearson

Sarah Boslaugh, Statistics in a Nutshell , O'Reilly Media, Inc Novembre 2012

Audet, Boucher, Caumartin, Skeene ; Probabilités et Statistique ; Gaëtan Morin éditeur

John Buglear , Practical Statistics, Kogan Page, novembre 2013

Sarah Stowell , Using R for Statistics, Apress, juin 2014

Practical Statistics for Data Scientists: 50 Essential Concepts Paperback – May 28 2017 by Peter Bruce (Author), Andrew Bruce (Author)