

Sigle : INF1473 Gr. 01

Titre : Entreposage et prospection de données

Session : Automne 2024 Horaires et local

Professeur : Tajeuna, Etienne

1. Description du cours paraissant à l'annuaire :

Objectifs

Au terme de ce cours, l'étudiant.e. saura construire un entrepôt de données, et appliquer des techniques de traitement de données, de recherche et d'analyse de l'information dans un contexte d'intelligence d'affaires (*Business Intelligence*).

Contenu

Entreposage de données : étapes de construction d'un entrepôt de données (acquisition, stockage, traitement et accès), modélisation multidimensionnelle des données et cubes de données, techniques OLAP, types d'architectures des entrepôts de données, optimisation des performances, produits et applications. Prospection de données : étapes de découverte de connaissances (prétraitement, prospection de données et interprétation des résultats), techniques de classification (arbres de décision, etc.), techniques de regroupement (treillis de concepts, classification hiérarchique), règles d'association et mesures de qualité, techniques statistiques élémentaires, produits et applications. Ce cours comporte des séances obligatoires de travaux dirigés (TD).

Descriptif – Annuaire

2. Objectifs spécifiques du cours :

- Familiariser l'étudiant(e) avec les techniques *Business Intelligence*.
- Présenter les principales techniques d'entreposage, prétraitement et prospection de données.
- Illustrer ces techniques à l'aide de *RapidMiner* pour le prétraitement et la prospection de données et du module SSAS (*SQL Server Analysis Services*) de Microsoft pour l'entreposage et la fouille de données.
- Permettre à l'étudiant(e) de mettre en pratique toutes les connaissances et techniques acquises durant le cours via des exercices et un projet de session portant sur le prétraitement, la prospection et l'entreposage de données.

3. Stratégies pédagogiques :

Les formules pédagogiques suivantes seront utilisées :

Logistique du cours

- Accès en laboratoire à des postes de travail sur lesquels le client SQL 2019 est installé.
- Accès à Moodle sur le Web pour la récupération des notes de cours, des énoncés de travaux, des consignes spécifiques, des soumissions des travaux et des résultats d'évaluation.

Plan synthétisé du cours

Les thèmes suivants seront étudiés :

- Introduction à la veille économique (*Business intelligence*) et à la prospection de données (*Data mining*)
- Techniques de prétraitement de données : 1- *Épuration*. 2 - *Intégration et transformation*. 3 - *Sélection et réduction*
- Entreposage de données : raison d'être et concepts
- Étapes de construction d'un entrepôt de données
- Modélisation multidimensionnelle
- Stratégies de conception
- Techniques OLAP
- Architectures des entrepôts de données
- Optimisation des performances

- Applications et produits
- Découverte de connaissances : raison d'être et définitions
- Étapes du processus de découverte de connaissances
- Survol des techniques de prospection de données et des applications courantes
- Classification, régression et prédiction : 1- Définitions, principaux thèmes. 2 - Comparaison entre la classification, la régression et la prédiction. 3 - Arbres de décision et règles de classification. 4 - Réseaux bayésiens. 5 - Approches statistiques de prédiction (modèles de régression)
- Regroupement : méthodes hiérarchiques comme K-Means
- Règles d'association et mesures de qualité

NOTE : Toutes les parties du cours seront illustrées avec les outils Business Intelligence de SQL Server 2019 et/ou la plateforme RapidMiner.

4. Heures de disponibilité ou modalités pour rendez-vous :

Disponible les mardi et mercredi matin entre 09h30 et 12h00 et le jeudi soir de 13h00 à 15h00.

Courriel : etienne.gael.tajeuna@uqo.ca

5. Plan détaillé du cours sur 15 semaines :

Semaine	Thèmes	Dates
1	Chap. 1- Business Intelligence (Intelligence d'affaires) Introduction	03 sept. 2024
2	Chap. 2- Techniques d'extraction, transformation et chargement de données Travaux dirigés 1 : ETL - Jeudi 12 sept. 2024	10 sept. 2024
3	Chap. 3- Modélisation multidimensionnelle et stratégies de conception Travaux dirigés 2 : Modélisation de données - Jeudi 19 sept. 2024	17 sept. 2024
4	Chap. 4- Construction d'un entrepôt de données (<i>Data Warehousing</i>) Travaux dirigés 3 : Chargement et manipulation des entrepôts de données - Jeudi 26 sept. 2024	24 sept. 2024
5	Chap. 5- Techniques OLAP et optimisation Projet P1 : Conception et manipulation d'un entrepôt de données	01 oct. 2024
6	Révision et discussion générale sur le cours Préparation pour l'examen de mi-session	08 oct. 2024
7	Semaine d'études	15 oct. 2024
8	Examen de mi-session Remise projet 1 : Vendredi 20 octobre 2024	22 oct. 2024
9	Chap. 6- Concepts de bases en prospection des données Prétraitement de données (Épuration, intégration, transformation, sélection et réduction de données). Techniques de réduction de la dimension (ACP, TSNE, UMAP). Techniques de sélection d'attributs Travaux dirigés 4 : Exploration et prétraitement des données - Jeudi 31 oct. 2024	29 oct. 2024
10	Chap. 7- Règles d'association. Introduction aux données transactionnelles. Principe de l'Algorithme Apriori. Recherche et extraction des règles d'association. Travaux dirigés 5 : Analyse d'association - Jeudi 07 nov. 2024	05 nov. 2024

11	<p>Chap. 8- Apprentissage non-supervisé</p> <p>Regroupement (clustering) par partition. Regroupement hiérarchique. Regroupement basé sur la densité. Validation des résultats de regroupement</p> <p>Travaux dirigés 6 : Méthodes de regroupements - Jeudi 14 nov. 2024</p> <p>Projet P2 : Prétraitement des données et regroupement</p>	12 nov. 2024
12	<p>Chap. 9- Apprentissage supervisé</p> <p>Classification et régression des données</p> <p>Travaux dirigés 7 : Méthodes de classification et régression - Jeudi 21 nov. 2024</p> <p>Remise projet 2 : Dimanche 24 nov. 2024</p>	19 nov. 2024
13	<p>Chap. 9- Apprentissage supervisé (suite et fin)</p> <p>Projet P3 : Classification et régression</p>	26 nov. 2024
14	<p>Révision et discussion générale sur le cours</p> <p>Préparation pour l'examen final</p>	03 déc. 2024
15	<p>Examen final</p> <p>Remise projet P3 : Dimanche 15 déc. 2024</p>	10 déc. 2024

6. Évaluation du cours :

L'étudiant(e) dans ce cours sera évalué(e) par les examens de mi-session et final, ainsi que par des travaux pratiques et sa présence aux séances de travaux dirigés. La pondération de la note finale sera comme suit :

- Examen de mi-session : **21 %**
- Examen final : **35 %**
- Projets (03): **30 % (10 % x 3)**
- Travaux dirigés (07 TDs): **14 % (2 % x 7)**

Les travaux pratiques comprendront les volets suivants : entreposage de données (01 TP), prétraitement des données (01 TP), regroupement de données (01 TP) et la prédiction (01 TP).

Toute note de moins de **50 % du cumul des deux examens** entraîne systématiquement un échec au cours. Les travaux pratiques se feront par des équipes de deux à trois étudiants et le libellé du TP **X** (= 1 ou 2) par l'équipe **N** doit être **INF1473-TPX-EquipeN**. La pénalité de retard pour la remise d'un travail est de **2 points** par jour (y compris les jours fériés et les fins de semaine).

Les présences aux séances de travaux dirigés seront considérées. Un(e) étudiant(e) qui s'absente souvent et de manière injustifiée aura une diminution de la note finale d'un maximum de **5 points**.

Des consignes sur l'échéancier et la réalisation des travaux pratiques seront précisées.

Des consultations de groupes seront organisées sur rendez-vous afin de guider et d'orienter les étudiant(e)s dans la réalisation de leurs travaux.

7. Politiques départementales et institutionnelles :

- Politique du département d'informatique et d'ingénierie relative à la tenue des examens
- Note sur le plagiat et sur la fraude
- Politique relative à la qualité de l'expression française écrite chez les étudiants et les étudiantes de premier cycle à l'UQO
- Absence aux examens : cadre de gestion, demande de reprise d'examen (formulaire)

Tolérance **ZÉRO** en matière de violence à caractère sexuel.

Le Bureau d'intervention et de prévention en matière de harcèlement (BIPH) a pour mission d'accueillir, soutenir et guider toute personne vivant une situation de harcèlement, de discrimination ou de violence à caractère sexuel. Le BIPH oriente ses actions afin de prévenir les violences à caractère sexuel pour que nous puissions étudier, travailler et s'épanouir dans un milieu sain et sécuritaire.

Vous vivez ou êtes une personne témoin d'une situation de violence à caractère sexuel ? Vous êtes une personne membre de la communauté étudiante ou une personne membre du personnel, autant à Gatineau qu'à Ripon et St-Jérôme, l'équipe du BIPH est là pour vous, sans jugement et en toute confidentialité.

Ensemble, participons à une culture de respect.

Pour de plus amples renseignements consultez [UQO.ca/biph](https://uqo.ca/biph) ou écrivez-nous au Biph@uqo.ca

8. Principales références :

Notes de cours disponibles sur Moodle (principale référence)

1. Bertrand Burquier. *Business intelligence avec SQL Server 2008*. Mise en œuvre d'un projet décisionnel, Dunod, 2009.
2. Sébastien Fantini. *Business Intelligence avec SQL Server 2019 et Power BI - Maîtrisez les concepts et réalisez un système décisionnel*, ENI, Mars 2020.
3. Matteo Golfarelli & Stefano Rizzi. *Data Warehouse Design: Modern Principles and Methodologies*, McGraw-Hill, 2009.
4. Kellyn Gorman, Allan Hirt, Dave Noderer, Mitchell Pearson, James Rowland-Jones, Dustin Ryan, Arun Sirpal, Buck Woody. *Introducing Microsoft SQL Server 2019: Reliability, scalability, and security both on premises and in the cloud*, Packt Publishing, 2020.
5. Jiawei Han, Jian Pei & Hanghang Tong. *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 4th edition, 2023.
6. Steven Hughes and Adam Jorgensen. *Hands-On SQL Server 2019 Analysis Services: Design and query tabular and multi-dimensional models using Microsoft's SQL Server Analysis Services*, Packt Publishing, 2020.
7. Bill H. Inmon. *Building the Data Warehouse*, John Wiley, 3^e édition, 2002.
8. Ralph Kimball & Margy Ross. *Guide pratique de modélisation dimensionnelle*, Vuibert informatique, Paris, 2002.
9. Ralph Kimball & Margy Ross. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, ~~Inde~~, Wiley, 2013
10. Stéphane Tufféry. *Data mining et statistique décisionnelle - L'intelligence des données*, éditions TECHNIP, 2012.
11. Alejandro Vaisman & Esteban Zimányi. *Data Warehouse Systems - Design and Implementation*, Springer, 2014.
12. Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher & J. Pal. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th Edition, Morgan Kaufmann, 2017.
13. Documentation sur RapidMiner Studio.
<https://docs.rapidminer.com/latest/studio/getting-started/> Page consultée le 5 décembre 2023
14. Module Turbo Prep de RapidMiner Studio
<https://docs.rapidminer.com/latest/studio/guided/turbo-prep/> Page consultée le 5 décembre 2023

9. Page Web du cours :

<https://moodle.uqo.ca>