# Fouille de Données et Apprentissage Automatique

TEM2004-NAIN_

Domaine : **Sciences** Grade académique : **Master en architecture des systèmes informatiques - Namur**
Niveau de certification : **7** Cycle : **Deuxième cycle** Année académique : **2026-2027** ( Q1 )

| CRÉDITS ECTS | VOLUME HORAIRE |
|---|---|
| **5** | **60 h** |
| PRÉREQUIS<br>*Aucun* | CO-REQUIS<br>*Aucun* |

| IMPLANTATION | DIRECTEUR DE DÉPARTEMENT DU DOMAINE | RESPONSABLE DE L'UE | ENSEIGNANTS INTERVENANTS |
|---|---|---|---|
| Département technique de Namur | Willem Kuypers | Laurent PHILIPPE | Laurent PHILIPPE<br>Thomas REMY |

PROGRAMME D'ÉTUDES

| Programme | Bloc | Caractère |
|---|---|---|
| IE-Master en architecture des systèmes informatiques | 2 | Obligatoire |

## Partie 1 Contribution de l'UE au profil d'enseignement du programme et acquis d'apprentissage spécifiques (AAS) sanctionnés par l'évaluation

### CONTEXTUALISATION

UE prérequise : aucune

Cette UE contribue à l'atteinte des acquis d'apprentissage terminaux suivants :

• Résoudre selon une approche analytique et systémique des problèmes liés à des situations nouvelles ou présentant un certain degré d'incertitude

– Identifier, traiter et synthétiser les données pertinentes

– Rechercher les ressources nécessaires

– Transposer les résultats des concepts à la problématique ou situation traitée

– Argumenter les choix proposés

• Communiquer dans un cadre professionnel avec les collaborateurs, les clients

– Rédiger des rapports, des cahiers des charges, fiches techniques ou manuels

– S'exprimer de manière adaptée en fonction du public

– Utiliser efficacement les outils de communication adéquats

• Agir de façon réflexive et autonome, en équipe, en partenariat

– Organiser son temps, respecter les délais

– Acquérir une capacité d'autocritique

– Actualiser ses connaissances et compétences

• Innover, concevoir ou optimiser un système

– Intégrer l'ensemble des composantes d'un système à partir des résultats d'analyse

– Modéliser, dimensionner et optimiser des systèmes

• Gérer les ressources techniques et financières de systèmes

– Estimer les coûts, la rentabilité d'un projet, établir un budget

– Evaluer les processus et les résultats

– Proposer et appliquer les actions correctives

• Utiliser des procédures et des outils

– Exploiter les outils appropriés afin de résoudre une tâche spécifique

– Effectuer des tests, des contrôles et des mesures

– Effectuer des paramétrages, des réglages

### ACQUIS D'APPRENTISSAGE SPÉCIFIQUES (AAS)

**Théorie et applications**

L'étudiant sera capable au terme de l'activité d'apprentissage de :

• situer la thématique de la fouille de données dans le contexte global (tant historique que fonctionnel) de la Business Intelligence (BI) et de son implémentation en entreprise

• expliquer (dans le détail, en référence à la théorie) et interpréter les différents paramètres d'un modèle d'apprentissage automatique mis au point dans un environnement de prototypage professionnel (Jupyter notebook - Python)

• faire du story telling correspondant à un cas d'usage précis (données + modèle)

• comparer des modèles proposés pour résoudre une problématique donnée suivant différentes dimensions : complexité de mise au point, de configuration, efficacité, robustesse, ...

• sélectionner un type de modèle et le mettre au point pour répondre à une problématique donnée - sélectionner, combiner et croiser des techniques et des concepts pour résoudre un problème complexe

• s'exprimer avec précision et rigueur tant à l'écrit qu'à l'oral pour délivrer un contenu scientifiquement viable

• préparer et tenir une keynote devant la classe sur un sujet d'actualité en rapport avec le contenu du cours

**Laboratoire**

L'étudiant sera capable au terme de l'activité d'apprentissage de :

• créer une présentation originale sur un sujet/une technique non traité(e) au cours, lié(e) au traitement automatique des données

• mettre au point un exercice illustrant concrètement la technique sous-jacente conformément aux consignes/contraintes (être exécutable en quelques minutes par l'ensemble de la classe) dans un environnement de prototypage professionnel

• chercher, sélectionner ou synthétiser/générer/construire un jeu de données pertinent

• présenter avec précision et rigueur la technique visée à l'ensemble de la classe à partir d'artefacts préparés

• mettre en rapport/expliciter les liens entre le sujet/la technique visée et les concepts de base de la partie Théorie & Applications

### OBJECTIFS DE DÉVELOPPEMENT DURABLE

Document annexe à la fiche UE accessible via un lien explicitant en quoi le programme d'études s'inscrit dans une politique de développement durable. Une ou plusieurs UE spécifique(s) pourraient être ciblée(s) dans cette annexe.

## Partie 2 Description de l'unité d'enseignement

### OBJECTIFS

*À définir*

### CONTENU

**Théorie et applications**

• Positionnement de la fouille de données dans le contexte de la Business Intelligence ou de l'Informatique Décisionnelle

• Rappels statistiques descriptives et inférentielles

• Concepts de sur et sous-apprentissage

• Régression linéaire simple et multiple

• Maximum de vraisemblance

• Évaluation des modèles - Matrice de confusion et autres caractérisations

• Courbe ROC - Lift

• Sélection de dimension - Principe d'Occam

• Techniques non supervisées

**Laboratoire**

Liste de sujets potentiellement à traiter comprend :

• Analyse en composantes principales

• Segmentation (clustering)

• Text Mining

• Modélisation de thématique (topic modelling)

• SVM (Support Vector Machine)

• Robot Process Automation (RPA)

• Réseaux de neurones

• Réseaux de neurones profonds

• ...

Dans les premières séances, l'enseignant présentera 3 sujets conformément à ce qui est attendu des groupes d'étudiants.

Les séances se poursuivront en autonomie pour la préparation des exposés, l'enseignant étant disponible et ressource durant ce temps.

### SOURCES ET RÉFÉRENCES

**Théorie et applications**

• Support de cours DATA MINING et DATA SCIENCE de Ricco Rakotomalala https://cours-machine-learning.blogspot.com/

• Apprentissage statistique — Réseaux de neurones, cartes topologiques, machines à vecteurs supports, G. Dreyfus, J.-M. Martinez, M. Samuelides, M. B. Gordon, F. Badran, S. Thiria (2008) Eyrolles

• Apprentissage artificiel - Concepts et algorithmes, A. Cornuéjols et L. Miclet (2010) Eyrolles

• Artificial Intelligence — A Modern Approach, Stuart Russel, Peter Norvig (2016) Pearson

• Machine Learning — A Probabilistic Perspective, Kevin P. Murphy (2012) The MIT Press

• Applied Predictive Modeling, Max Kuhn et Kjell Johnson (2013) Springer

**Laboratoire**

Les étudiants doivent trouver par eux-mêmes les sources originales et scientifiquement valides afin de présenter leur travail.

### SUPPORTS DE COURS

| Activité d'apprentissage | Support |
|---|---|
| **AAIN2004-A** — Théorie et applications | Les jeux de slides présentés ainsi que les énoncés et les correctifs des exercices dirigés sont disponibles sur la plateforme Moodle. |
| **AAIN2004-B** — Laboratoire | La liste des sujets potentiels ainsi que les exemples de présentations faites par l'enseignant sont disponibles sur la plateforme Moodle.. |

## Partie 3 Méthodologie mise en œuvre et éventuel découpage de l'UE

### DESCRIPTION DES ACTIVITÉS D'APPRENTISSAGE (AA) AU REGARD DE LEUR STATUT

| Activité d'apprentissage | Heures | Langue d'enseignement | Langue d'évaluation |
|---|---|---|---|
| **AAIN2004-A** — Théorie et applications | 45 h | français mais une compréhension à la lecture de l'anglais peut être nécessaire (données, métadonnées, textes d'explication référencés dans les énoncés ou le matériel de support) | français mais une compréhension à la lecture de l'anglais peut être nécessaire (données, métadonnées, textes d'explication référencés dans l'énoncé) |
| **AAIN2004-B** — Laboratoire | 15 h | français | français |

### MÉTHODES D'ENSEIGNEMENT ET D'APPRENTISSAGE

| Activité d'apprentissage | Méthodes d'enseignement et d'apprentissage |
|---|---|
| **AAIN2004-A** — Théorie et applications | `- Partie théorique ex-cathédra`<br>`- Exposés multimédias`<br>`- Exercices - Travaux dirigés : au travers des exercices proposés, les étudiants apprennent à utiliser les librairies de Data M`<br>`- Synthèses collectives` |
| **AAIN2004-B** — Laboratoire | • Exposés exemplatifs par le professeur<br>• Travaux de groupe en autonomie<br>• Exercice de recherche sur un sujet pointu, vulgarisation et présentation pédagogique à l'ensemble de la classe<br>• Dans une liste de sujets potentiels:<br>◦ Analyse en composantes principale<br>◦ Segmentation (clustering)<br>◦ Text Mining<br>◦ SVM<br>◦ ...<br>L'enseignant présentera, dans les premières séances, 3 sujets conformément à ce qui est attendu des groupes d'étudiants. Les séances se pours |

### DÉTERMINATION DE LA COHÉRENCE PÉDAGOGIQUE EN CAS DE REGROUPEMENT D'AA FAISANT L'OBJET D'UNE ÉVALUATION DISTINCTE

*À définir*

## Partie 4 Évaluation de l'unité d'enseignement

### STRUCTURE DU DISPOSITIF GÉNÉRAL D'ÉVALUATION (EN LIEN AVEC LE PROGRAMME D'ÉTUDES)

Évaluation intégrée : non

UE remédiable.

Évaluation : session de janvier (Q1) et seconde session (Q3).

Note de l'UE : moyenne pondérée des notes des activités d'apprentissage :

• AAIN2004-A – Théorie et applications : 75 %

• AAIN2004-B – Laboratoire : 25 %

### MÉTHODOLOGIE DE CONSTRUCTION DE LA NOTE FINALE DE L'UE

Se référer à l'évaluation des activités d'apprentissage.

### LA PONDÉRATION DE L'ÉVALUATION DE L'UE À DES FINS DE DÉLIBÉRATION LORS DU CALCUL DE LA MOYENNE

Au sein du programme d'études, lors du calcul de la moyenne, la pondération de l'évaluation de l'unité d'enseignement est effectuée sur base des crédits affectés à cette unité d'enseignement et évalués

### MODALITÉS D'ÉVALUATION PAR ACTIVITÉ D'APPRENTISSAGE

| Activité d'apprentissage | Modalités d'évaluation |
|---|---|
| **AAIN2004-A** — Théorie et applications | `- Q1 : examen écrit (4H) à livre ouvert`<br>`- Q3 : examen écrit (4H) à livre ouvert` |
| **AAIN2004-B** — Laboratoire | **SEPTEMBRE - JANVIER**<br>Examen oral hors session à la fin du quadri Q1.<br>**DEUXIÈME SESSION**<br>Examen Oral.<br>L'évaluation tiendra compte des dimensions suivantes (liste non exhaustive):<br>• la sélection du contenu (pertinence dans un contexte pédagogique,...)<br>• la compréhension de ce contenu<br>• la pédagogie de l'exposé (fil rouge, exercice et data set préparés)<br>• la gestion du temps<br>• l'équilibrage des ressources humaines lors de l'exposé<br>• l'autorégulation dynamique du groupe (y compris les réactions vis-à-vis des questions de la classe)<br>• l'homogénéité de l'exposé |

### MODALITÉS CONCRÈTES D'ÉVALUATION

Une mention sur la fiche informera les étudiants du fait que les précisions sur les épreuves seront disponibles sur Moodle au plus tard 6 semaines avant l'évaluation.