

**MINISTERE DE LA COMMUNAUTE FRANCAISE**  
**ADMINISTRATION GENERALE DE L'ENSEIGNEMENT**  
**ENSEIGNEMENT DE PROMOTION SOCIALE**

**DOSSIER PEDAGOGIQUE**

**UNITE D'ENSEIGNEMENT**

**STATISTIQUES ET DATA MINING**

**ENSEIGNEMENT SUPERIEUR DE TYPE COURT**  
**DOMAINE : SCIENCES ÉCONOMIQUES ET DE GESTION**

**CODE : 75 69 09 U32 D1**

**CODE DU DOMAINE DE FORMATION : 710**

**DOCUMENT DE REFERENCE INTER-RESEAUX**

**Approbation du Gouvernement de la Communauté française du 19 juillet 2019,  
sur avis conforme du Conseil général**

# STATISTIQUES ET DATA MINING

## ENSEIGNEMENT SUPERIEUR DE TYPE COURT

### 1. FINALITES DE L'UNITE D'ENSEIGNEMENT

#### 1.1. Finalités générales

Conformément à l'article 7 du décret de la Communauté française du 16 avril 1991 organisant l'enseignement de promotion sociale, cette unité d'enseignement doit :

- ◆ concourir à l'épanouissement individuel en promouvant une meilleure insertion professionnelle, sociale, culturelle et scolaire ;
- ◆ répondre aux besoins et demandes en formation émanant des entreprises, des administrations, de l'enseignement et d'une manière générale des milieux socio-économiques et culturels.

#### 1.2. Finalités particulières

Cette unité d'enseignement vise à permettre à l'étudiant :

- ◆ de s'approprier des connaissances générales en Data Mining en vue de traiter des données économiques, financières, commerciales...;
- ◆ de développer son sens critique dans l'interprétation de résultats obtenus lors de l'application de différentes méthodes d'analyses statistiques sur divers types de données ;
- ◆ d'utiliser des ressources logicielles existantes permettant le traitement statistique de données pour décrire des interrelations entre des variables choisies et les illustrer graphiquement ;
- ◆ développer des modèles descriptifs et prédictifs d'analyse.

### 2. CAPACITES PREALABLES REQUISES

#### 2.1. Capacités

*face à une situation issue de la vie professionnelle courante,  
à partir de documents dont certains rédigés en langue anglaise,  
dans le respect des règles et usages de la langue française,*

- ◆ présenter une problématique abordée dans le texte et poser une question de départ,
- ◆ proposer une hypothèse de travail dans au moins un des domaines suivants :
  - économie,
  - informatique,
  - technique,
  - secteur public ;
- ◆ répondre de manière claire, concise et argumentée aux questions qui lui seront posées ;
- ◆ mettre en œuvre de manière adéquate les fonctionnalités d'une base de données tabulaire ;
- ◆ appréhender la gestion de bases de données et les technologies associées.

## 2.2. Titre pouvant en tenir lieu

Grade de master ou de bachelier de type court défini dans la liste tenue et mise à jour par l'Académie de Recherche et d'Enseignement Supérieur (ARES).

## 3. ACQUIS D'APPRENTISSAGE

**Pour atteindre le seuil de réussite, l'étudiant sera capable :**

*face à au moins une situation se prêtant à la mise en œuvre de méthodes de traitement de données opérationnelles par l'outil statistique,*

- ◆ de formuler des hypothèses nulles et alternatives et de les tester en recourant à des techniques appropriées ;
- ◆ d'interpréter les résultats de ces tests d'hypothèses et d'en déduire leurs conséquences ;
- ◆ de déterminer au moins une méthode appropriée d'analyse statistique en fonction de la problématique à analyser (multi-variée) ;
- ◆ de déterminer et d'appliquer, à l'aide des outils informatiques appropriés et disponibles, les principales techniques descriptives et prédictives du Data Mining en justifiant les choix.

**Pour la détermination du degré de maîtrise, il sera tenu compte des critères suivants :**

- ◆ niveau de précision : la clarté, la concision, la rigueur au niveau de la terminologie, des concepts et des techniques/principes/modèles,
- ◆ niveau de cohérence : la capacité à établir avec pertinence une majorité de liens logiques pour former un ensemble organisé,
- ◆ niveau d'intégration : la capacité à s'approprier des notions, concepts, techniques et démarches en les intégrant dans son analyse, son argumentation, sa pratique ou la recherche de solutions,
- ◆ niveau d'autonomie : la capacité de faire preuve d'initiatives démontrant une réflexion personnelle basée sur une exploitation des ressources et des idées en interdépendance avec son environnement.

## 4. PROGRAMME

L'étudiant sera capable :

*face à au moins une situation se prêtant à la mise en œuvre de méthodes de traitement de données opérationnelles par l'outil statistique,*

### **En statistiques inférentielles**

- ◆ d'illustrer les notions de probabilité, de distribution d'une variable probabilisée et d'interpréter les valeurs en fonction des paramètres fondamentaux relatifs à une distribution de données ;
- ◆ d'expliquer :
  - les notions de significativité en rapport aux principaux tests d'hypothèses (sur la moyenne et la dispersion),
  - l'interprétation de leurs résultats (seuil de significativité et intervalle de confiance) ;

- ◆ d'expliquer les différents types de liaison puis la corrélation (linéaire ou non) et la régression linéaire (simple et multiple) ;
- ◆ d'appréhender et d'analyser une démarche de résolution de problème et de la justifier en fonction des conditions initiales de la problématique à analyser, notamment :
  - d'analyser le problème posé et de formuler des hypothèses (nulles ou alternatives),
  - de déterminer les techniques appropriées pour les tester et d'interpréter les résultats de ces tests d'hypothèses et leurs conséquences ainsi qu'en déduire des conclusions,
  - de déterminer les méthodes les plus appropriées d'analyse statistique inférentielle (multi-variée).

### **En Data Mining**

- ◆ d'en lister les utilités principales (classer, estimer, segmenter, associer, prévoir ...) ;
- ◆ d'expliquer les principales techniques descriptives et prédictives du Data Mining, et d'en préciser les algorithmes et conditions d'utilisation, à savoir :
  - au niveau descriptif : analyse factorielle et analyse de correspondances, recherche d'associations, clustering dynamique...,
  - au niveau prédictif : analyse discriminante, régressions simples et multiples, analyses de variance ANOVA et MANOVA...,
- ◆ d'appliquer ces approches avec un logiciel adapté (SPSS, SAS Base...) et d'établir des relations avec d'autres packages dans des domaines spécifiques (CRM, ERP, WEB Analytics ...) ;
- ◆ d'expliquer les apports de la méthode du Data Mining et de la comparer avec l'analyse statistique classique.

## **5. CHARGE(S) DE COURS**

Le chargé de cours sera un enseignant ou un expert.

L'expert devra justifier de compétences particulières issues d'une expérience professionnelle actualisée en relation avec le programme du présent dossier pédagogique.

## **6. CONSTITUTION DES GROUPES OU REGROUPEMENT**

Il est recommandé de ne pas dépasser plus d'un étudiant par poste de travail.

## **7. HORAIRE MINIMUM DE L'UNITE D'ENSEIGNEMENT**

<b>7.1. Dénomination des cours</b>	<b><u>Classement</u></b>	<b><u>Code U</u></b>	<b><u>Nombre de périodes</u></b>
Statistiques inférentielles	CT	B	28
Data Mining : laboratoire	CT	S	36
<b>7.2. Part d'autonomie</b>		P	16
<b>Total des périodes</b>			<b>80</b>