

# Entreposage et fouille de données

## Présentation

### Public, conditions d'accès et prérequis

être admis en M2 du master STIC, mention statistique ou être agréé.

**La demande d'agrément est à faire uniquement pour les auditeurs qui souhaitent suivre STA211 en présentiel au semestre 1.** Pour obtenir l'agrément, les auditeurs adresseront, **à partir du 25 Août**, par courrier électronique à l'enseignant responsable : **ndeye.niang\_keita@cnam.fr**, un CV détaillé et une lettre de motivation indiquant les raisons de la demande et le projet pédagogique dans lequel elle s'inscrit. Une réponse sera donnée dans un délai d'une dizaine de jours.

Niveau requis : NFA008 (bases de données) et STA101 (analyse des données, méthodes descriptives). **Ces prérequis sont indispensables pour obtenir l'agrément. Vérifier avant d'envoyer une demande d'agrément.**

### L'avis des auditeurs

Les dernières réponses à l'enquête d'appréciation pour cet enseignement :

[Fiche synthétique au format PDF](#)

### Présence et réussite aux examens

Pour l'année universitaire 2021-2022 :

Nombre d'inscrits : 105

Taux de présence à l'évaluation : 58%

Taux de réussite à l'évaluation : 43%

### Mots-clés

[Analyse statistique des données](#)

[Statistique décisionnelle](#)

[Aide à la décision](#)

[Data mining](#)

[Base de données](#)

[fouille de données](#)

## Programme

### Contenu

#### Modèles prévisionnels et systèmes de gestion de l'entreprise

- structures spécifiques des bases de données de Data warehouse (star schema)
- OLAP

#### Méthodologies générales

- Méthodologies de Data Mining

#### Pré-traitement des données

- Analyses de la qualité des données,
- Techniques d'appréhension des valeurs manquantes ou aberrantes
- Techniques de construction de bases de travail (agrégations, etc. . . )

### Données et techniques de fouille

Méthodes non supervisées :

- Cartes de Kohonen, Règles d'association

Méthodes supervisées :

- Rappels de théorie de l'apprentissage
- Arbres de décision, forêts aléatoires, Réseaux de neurones, deep learning
- Méta-algorithmes :
- boosting, bagging

Fouille dans de nouveaux types de données et méthodes associées :

- Données textuelles - Données multivues - Images et Multimedia

Outils :

- Environnements freeware : R, Python
- Outils spécifiques : SAS-EM, SPAD
- Data Mining et bases de données : OLAP Business Object

## Modalité d'évaluation

Projet(s)  
Mémoire

## Bibliographie

M.BARDOS : Analyse discriminante (Dunod, 2001)  
 G.SAPORTA : Probabilités, analyse des données et statistique (Technip, 2006)  
 S.TUFFERY : Data mining et statistique décisionnelle (Technip, 2005)  
 S.TUFFERY : Etude de cas en statistique décisionnelle (Technip, 2009)  
 T.HASTIE, J.FRIEDMAN, F.TIBSHIRANI : Elements of Statistical Learning (Springer, 2009)  
 G.GOVAERT (ed) : Analyse des données (Hermes,2003)  
 L. LEBART, A. MORINEAU, M. PIRON : Statistique exploratoire multidimensionnelle (1995)  
 J.P.NAKACHE, J.CONFAIS : Statistique explicative appliquée (Technip, 2003)  
 James, Witten, Hastie, & Tibshirani : An Introduction to Statistical Learning (2013) Téléchargeable  
 ici:[http://web.stanford.edu/~hastie/local.ftp/Springer/ISLR\\_print1.pdf](http://web.stanford.edu/~hastie/local.ftp/Springer/ISLR_print1.pdf)

Parcours

## Cette UE apparaît dans les diplômes et certificats suivants

Chargement du résultat...



| Intitulé de la formation | Type | Modalité(s) | Lieu(x) | Intitulé de la formation | Type | Modalité(s) | Lieu(x) |
|--------------------------|------|-------------|---------|--------------------------|------|-------------|---------|
| Intitulé de la formation | Type | Lieu(x)     | Lieu(x) | Intitulé de la formation | Type | Lieu(x)     | Lieu(x) |
| Intitulé de la formation | Type | Lieu(x)     | Lieu(x) |                          |      |             |         |
| Intitulé de la formation | Type | Lieu(x)     | Lieu(x) |                          |      |             |         |
| Intitulé de la formation | Type | Lieu(x)     | Lieu(x) |                          |      |             |         |
| Intitulé de la formation | Type | Lieu(x)     | Lieu(x) |                          |      |             |         |

Informations pratiques

## Contact

## Voir le calendrier, le tarif, les conditions d'accessibilité et les modalités d'inscription dans le(s) centre(s) d'enseignement qui propose(nt) cette formation.

### UE

#### [Paris](#)

##### Centre Cnam Paris

2024-2025 1er semestre : Formation en présentiel soir ou samedi  
2024-2025 2nd semestre : Formation ouverte et à distance (FOAD)  
2025-2026 1er semestre : Formation en présentiel soir ou samedi  
2025-2026 2nd semestre : Formation ouverte et à distance (FOAD)  
2026-2027 1er semestre : Formation en présentiel soir ou samedi  
2026-2027 2nd semestre : Formation ouverte et à distance (FOAD)

#### Comment est organisée cette formation ?

```
/**/ details.orga-container { display: list-item; } details.orga-container summary { display: list-item; margin: 0.5em; color: #c1002a; font-weight: bold; cursor: pointer; } details.orga-container .orga-subtitle { margin-left: 1em; } details.orga-container .orga-list { margin-left: 1em; margin-bottom: 1em; } details.orga-container h3 { font-size: 1.2em; } /**/
```

2024-2025 1er semestre : Formation en présentiel soir ou samedi

#### Dates importantes

Période des séances du 16/09/2024 au 18/01/2025  
Période d'inscription : du 10/06/2024 à 10:00 au 18/10/2024 à 23:59  
Date de 1ère session d'examen : la date sera publiée sur le site du centre ou l'ENF  
Date de 2ème session d'examen : la date sera publiée sur le site du centre ou l'ENF

#### Précision sur la modalité pédagogique

Une formation en présentiel est dispensée dans un lieu identifié (salle, amphi ...) selon un planning défini (date et horaire).

2024-2025 2nd semestre : Formation ouverte et à distance

#### Dates importantes

Période des séances du 03/02/2025 au 07/06/2025  
Période d'inscription : du 10/06/2024 à 10:00 au 14/03/2025 à 23:59  
Date de 1ère session d'examen : la date sera publiée sur le site du centre ou l'ENF  
Date de 2ème session d'examen : la date sera publiée sur le site du centre ou l'ENF

#### Précision sur la modalité pédagogique

Une formation ouverte et à distance (FOAD) est une formation dispensée 100% à distance, qui peut être suivie librement, à son rythme.  
Regroupements physiques facultatifs : Aucun

#### Organisation du déploiement de l'unité

Nombre d'élèves maximum à distance par classe : 80  
Nombre d'heures d'enseignement par élève : 77  
Délai maximum de réponse à une sollicitation : sous **96** heures (*Jours ouvrés*)

## Modes d'animation de la formation

Forum  
Organisation d'une séance de démarrage  
Evaluation de la satisfaction  
Hot line technique

## Ressources mises à disposition sur l'Espace Numérique de Formation

Documents de cours  
Enregistrement de cours  
Documents d'exercices, études de cas ou autres activités pédagogiques  
tests de connaissance QCM

## Modalité de contrôle de l'acquisition des compétences et des connaissances (validation de l'UE)

Projet(s) individuel(s)

### Code UE : STA211

Cours

9 crédits

Volume horaire de référence  
(+ ou - 10%) : **70 heures**

#### **Responsable(s)**

Ndeye NIANG KEITA

Vincent AUDIGIER