



Spécialisation

« Science des données »



Responsable de la spécialisation :
François Husson : francois.husson@institut-agro.fr

Master Mathématiques Appliquées, Statistique

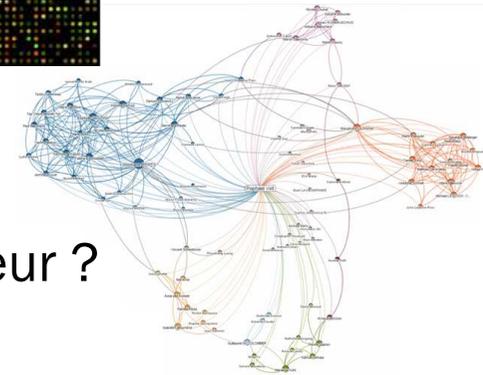
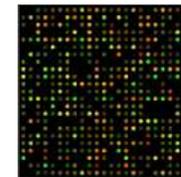
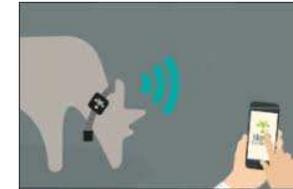
Parcours « Sciences des données
biologiques »



Responsable du master :
David Causeur : david.causeur@institut-agro.fr

Exemples de questions en science des données en production, alimentation, écologie ...

- Comment suivre en temps réel des exploitations équipées de capteurs pour conduire un élevage ou une vigne ?
- Comment analyser des données de drones et des images satellites pour raisonner la fertilisation ?
- Y a-t'il un lien entre contaminants d'origine alimentaire et risque de démence chez le sujet âgé ?
- Peut-on identifier les gènes responsables d'une maladie ?
- Comment utiliser la spectrométrie et l'imagerie en nutrition ?
- Comment représenter et exploiter les données d'un réseau social pour mieux comprendre le comportement du consommateur ?



Objectifs de la formation

Méthodes pour l'analyse de données

- Problématiques émergentes (données temporelles, hétérogènes, volumineuses, grande dimension, ...)
- Science des données, intelligence artificielle, informatique
- Contenu de formation diversifié (Bayésien, données textuelles, chimiométrie, données de perception ...) défini par les besoins

Expertise métier

- Enseignement orienté vers la confrontation aux problèmes

Préparation à l'insertion professionnelle

- Rencontres avec des professionnels
- Intégration dans un réseau d'employeurs

Contenu de la formation M1

- Semestre 2 (S8) – UE obligatoires (1UE = 50h)

Analyse de données (agronomes)

Planification expérimentale – Sensométrie ou Analyse de données génomiques

Programmation scientifique

Gestion des données - Statistique et aide à la décision

UE non obligatoires à choisir selon votre projet professionnel
(alimentation, agro-écologie, environnement, végétal, élevage, etc.)

Contenu de la formation M2

Semestre 1

Séquence 1 : méthodes (sept.-nov.)

- UE1. Réduction de la complexité
 - Analyse factorielle
 - Visualisation de données massives et hétérogènes
- UE2. Statistique pour données biologiques
 - Données expérimentales
 - Apprentissage de données biologiques
 - Statistique Bayésienne (écologie)
OU Analyse de données de perception
- UE3. Apprentissage statistique
 - Classification non-supervisée
 - Machine learning
- UE4. Méthodes informatiques
 - Big data avec R
 - Gestion de données massives
- UE5. Langues

Séquence 2 : projet de master (déc.-fév.)

Semestre 2 : Stage

Projet d'ingénieur

Objectif : Intégrer ses connaissances/compétences au service d'une problématique professionnelle

→ En 2022-2023

- Améliorer la reconnaissance d'espèces sur des images avec la classification taxonomique ([TerrOïko](#))
- Facial Aging Digital Evaluation : Apprentissage à partir de données dermatologiques ([Chanel](#))
- Suivi acoustique d'espèces sentinelles et parapluie : le cas des oiseaux crépusculaires et nocturnes de la réserve naturelle des nouragues ([Muséum d'histoire naturelle](#))
- Estimation de la composition en volatils à partir de capteurs - E-nose pour une application médicale ([Institut de Recherche Dupuy de Lôme - Centre Hospitalier Kerpape](#))
- Analyse des séquences de marche pour personnaliser la rééducation des patients atteints d'artériopathie oblitérante des membres inférieurs ([CHU Rennes - ENS Rennes](#))
- Cell deconvolution : méthodes statistiques pour identifier les types de cellule dans des échantillons à partir du transcriptome médicale ([Institut Pasteur - Université Paris Cité](#))
- Interprétabilité du Machine Learning sur des images de télédétection pour prédire la teneur en carbone des sols ([Institut Agro - labo recherche spatialisation numérique](#))
- Recueil et analyse de données pour comprendre l'écholocation

Projet d'ingénieur

Objectif : Intégrer ses connaissances/compétences au service d'une problématique professionnelle

→ En 2021-2022

- Imagerie racinaire : Traitement et segmentation automatique d'images de racines (INRAe)
- Préférences d'habitat du Puffin de Scopoli au regard de ses déplacements (INSERM)
- Lien entre données explicites et données implicites (eye-tracking) (Renault)
- Etude de l'exposition des abeilles aux pesticides (ANSES)
- Identification de signal par analyse de variance fonctionnelle
- Impact des séquences de traitement sur la survie des femmes atteintes d'un cancer du sein (Institut Curie)

→ En 2020-2021

- Impact du stress thermique sur les performances zootechniques des vaches laitières (BCEL)
- Etude de l'abondance et de la distribution spatiale des requins peau bleu (protocole d'observation depuis un avion) (Inserm)
- Typologie des magasins du réseau Biocoop à partir de données d'achats (Biocoop)
- Prédiction du nombre de fois où un chat mâche une croquette à partir de fichiers sons (Mars)
- Utilisation du machine learning pour composer un produit bio en fonction de valeurs nutritionnelles (Alma food)



Stages

<http://datasciences.agrocampus-ouest.fr/stages/>

- **Réseau d'employeurs**

- Ingénierie biostatistique/bioinformatique : INSERM, ANSES, Institut Curie, Institut Pasteur, INRAé, ARVALIS, ...
- Etudes consommateurs, analyse sensorielle, santé, marketing, ...

- **Stages**

- **Recherche** : INRAé, Inserm, Université (Lyon, Rennes, Paris, Nancy), CIRAD, IRISA, Institut Pasteur, Muséum national d'histoire naturelle, EHESP
- **Ecologie** : TerrOïko, office français de la biodiversité, Centre d'Ecologie Fonctionnelle et Evolutive, Cirad
- **Alimentation** : Entremont, Lactalis, ITM alimentaire international, labo national boulangerie pâtisserie, Credoc
- **Cosmétique** : L'Oreal, Chanel, Givaudan, Yves Rocher
- **Sensoriel étude conso** : Techni' Sens, Eurosyn, Kanthar, Danone
- **Instituts d'études** : Kanthar health
- **Divers** : Sanofi, EP, Potloc, Crédit Agricole, Ubisoft

• Stages 2022-23

- **Inrae** : Deep learning sur la prévision de floraison des pommiers
- **Adventiel** : Monitoring bovin outdoor et remote sensing grâce à l'IA, afin de factualiser des indicateurs de bien-être animal, la traçabilité des déplacements en assurant l'optimisation du pâturage
- **INRAE** : Approche de science participative pour étudier l'effets des IAE sur les communautés de chiroptères en paysage viticole
- **Inserm** : Exposition aux contaminants d'origine alimentaire et risque de démence chez le sujet âgé
- **Centre d'Ecologie Fonctionnelle et Evolutive** : Développement de méthodes de Deep Learning pour détecter les oiseaux sur fond cryptique
- **Centre d'Ecologie Fonctionnelle et Evolutive** : Identification d'individus de lynx/zèbres et étude statistique de l'impact de l'identification des individus par deep learning sur l'estimation des effectifs
- **CIRAD** : Caractérisation de l'entomofaune des agrosystèmes cannières et étude de l'influence du climat et des éléments paysagers sur les lépidoptères foreurs de tige et leurs ennemis naturels dans Petite-Île à la Réunion
- **Muséum national d'histoire naturelle** : Création d'indicateurs nationaux de la migration des passereaux communs de France par l'analyse des données de baguage
- **Inrae**: Effets des obstacles à l'écoulement sur la composition taxonomique et fonctionnelle des espèces de poissons amphihalins vs. exotiques et envahissants dans le cadre complexe des réseaux dendritiques

Production

Ecologie,
environnement

Stages

• Stages 2022-23

- Cosmétique** {
 - **Chanel** : Prédiction de la performance cosmétique
- Santé** {
 - **Irset** : Comparaison de la durée d'hospitalisation de patients ayant une chirurgie de l'hyperplasie bénigne de la prostate alors qu'ils sont exposés à des antidépresseurs et/ou à des antiagrégants plaquettaires
 - **Iplesp** : Analyser un problème dans le cadre d'un projet ou d'une étude réalisée au sein d'une structure institutionnelle professionnelle
 - **Novartis** : Analyses exploratoires et visualisation de données de biomarqueurs dans les études cliniques
 - **EHESP** : La relation entre le statut économique des patients et la maladie de démence
 - **Sanofi** : Etablissement et comparaison des calculs de Proven Acceptable Range en univarié et multivarié dans le cadre des caractérisations de procédés ou de méthodes analytiques
- Alimentation** {
 - **Credoc** : Observation des comportements alimentaires et de leurs évolutions en France
 - **Urbanlab** : Analyses, interprétation et visualisation d'open data pour le projet DatAgora (métropole de Lyon)
 - **Markem-Image industries**: Application de Machine Learning IOT pour la maintenance prédictive
- Divers** {
 - **EP** : Concevoir, modéliser et développer des modèles de Data science pour analyser les données dans le domaine de l'habitat et de la transition énergétique
 - **Ubisoft** : Développer des process analytics et de visualisation Big Data adéquats et innovants pour aider à la mise en production d'un jeu vidéo « Game as a Service »

• Stages 2020-21

- Production**
 - **Abelio** : Construction d'un outil d'aide à la décision pour la surveillance de culture
 - **CATE** : Analyse de données microbiennes des sols en culture maraîchère
 - **Syngenta** : Développement d'une application Shiny permettant de vérifier l'adéquation entre recommandations du Site Production Research et ce qui est fait au champs
- Ecologie**
 - **Eaden** : Reconnaissance faciale et corporelle de vaches à partir d'analyses d'images
- Alimentation**
 - **TerrOïko** : Comparaison de données génétiques mesurées (biblio) et simulées (SimOïko) sur diverses espèces pour évaluer le réalisme de SimOïko pour estimer la biodiversité
 - **Entremont** : Développement d'outils de prédiction afin de pouvoir maîtriser la qualité finale des fromages
 - **LACTALIS** : Explication des variations de rendement protéique des fromages
- Santé**
 - **Inserm** : Exposition aux contaminants d'origine alimentaire et risque de démence chez le sujet âgé
 - **CRAN** : Machine learning en génomique pour la santé humaine
 - **LTSI** : Traitement d'images médicale cérébrale pour localiser des régions fonctionnelles
 - **SANOFI** : Application Shiny pour visualiser les analyses de données Biomarqueurs
 - **Inserm** : Analyse spatio-temporelle de la mortalité globale au cours de l'épidémie de Covid-19, avec un focus sur l'impact du contexte social de chaque zone géographique sur les disparités de surmortalité

Stages

• Stages 2020-21

- **Chanel** : Prédiction de la performance de protection solaire d'une crème solaire grâce à des données de formulation
- **Givaudan** : Détection de thèmes émotionnels à travers l'analyse sémantique de données textuelles conso
- **L'Oréal** : Analyse et modélisation de données autour de la coloration capillaire
- **L'Oreal** : Construction d'un outil Shiny pour l'analyse de données sensorielles
- **Eurosyn** : Profil idéal sur l'amorti de chaussures de running.
- **Techni'Sens** : Etude de prédictivité de l'achat - mettre en évidence l'importance de certaines questions dans la prédictivité des projections d'achat des consommateurs
- **Crédit Agricole** : Analyse de comportements bancaires des clients par Machine Learning pour optimiser le ciblage des clients les plus appétents au crédit à la consommation
- **Université Lyon** : Analyse longitudinale de l'évolution du rapport aux langues lors de l'apprentissage chez l'enfant bilingue (français - anglais)
- **Université Paris Diderot** : Etude de stratégies d'agrégation consensuelle en apprentissage statistique. Application au problème de classification supervisée des voyelles

Cosmétique

Etude conso

Divers

Data challenge

Une compétence remarquée et récompensée par 2 prix : meilleur projet et numérique responsable

18 équipes de Rennes 1, Rennes 2, ENSAI, Clermont, Nantes, Paris, ... et l'institut Agro

Prix du numérique responsable

[Groupama Loire Bretagne](#) (2000 €)



Prix du Meilleur projet [Enedis](#) (2000 €)

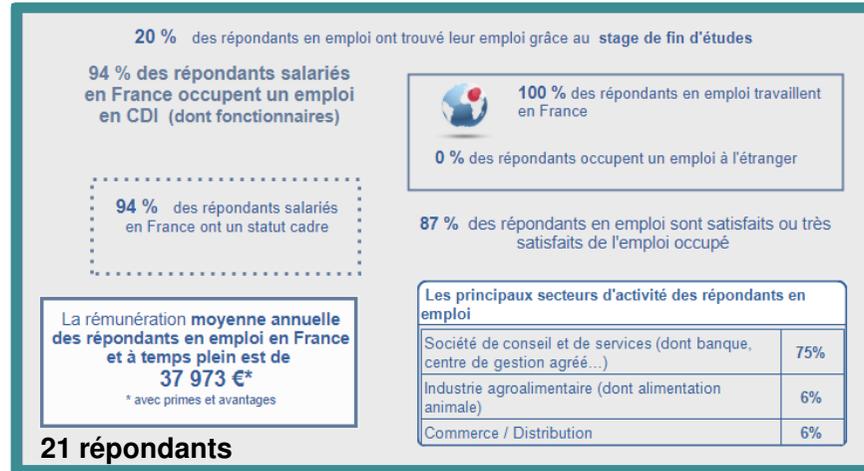
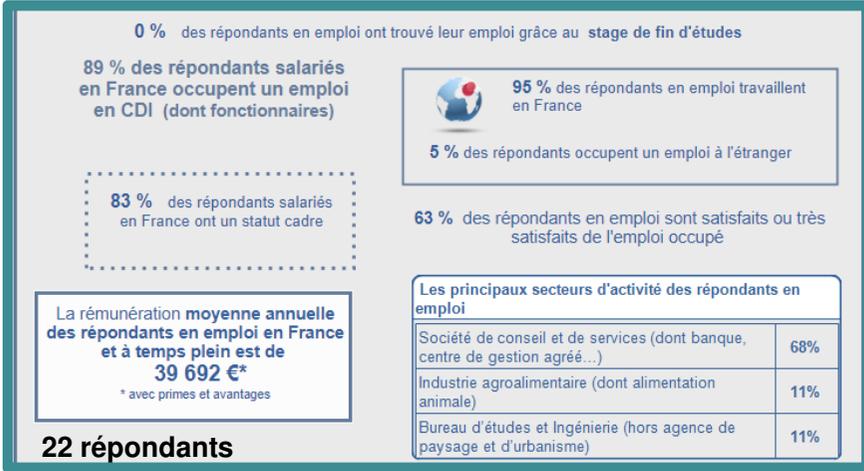


Insertion professionnelle

Agroalimentaire

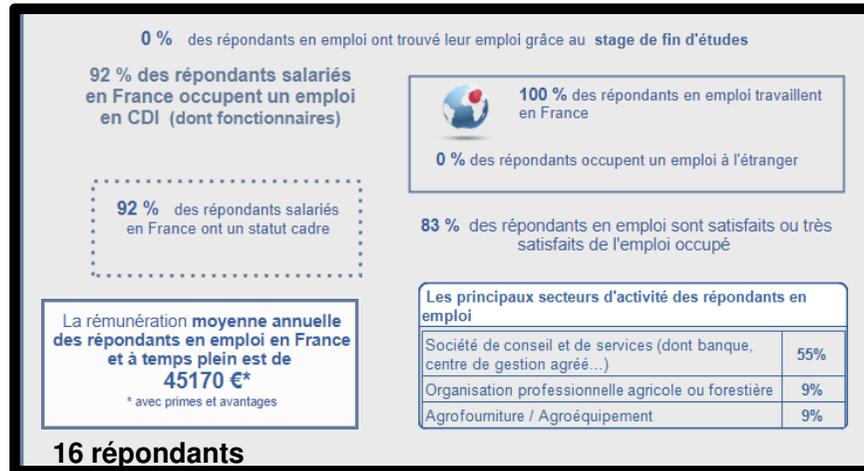
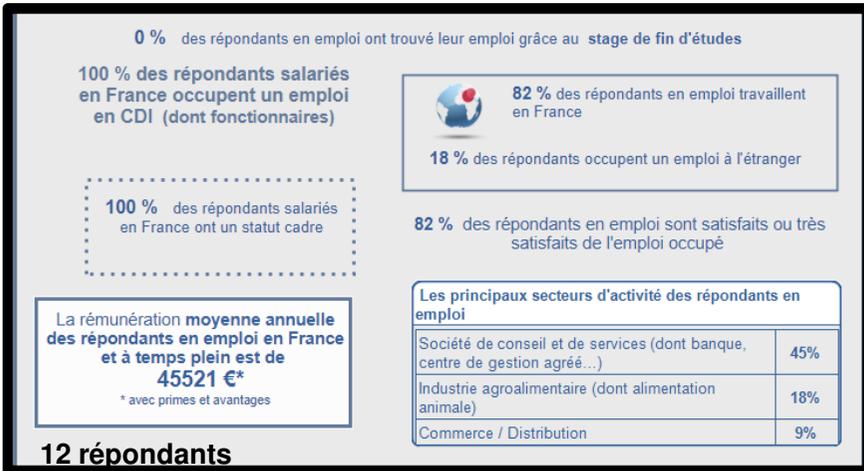
Agronome

Enquête à 12 mois



10 à 15% des étudiants font une thèse

Enquête à 5 ans



Commentaires des étudiants de la promo 2023-24

Les cours sont très intéressants

Top, enseignant au top et motivé

Conforme à mes attentes, très intéressant, stimulant et varié

La formation offrait une diversité de cours pertinents et de bonne qualité pour le domaine. Dans l'ensemble les cours sont bien équilibrés entre théorie et application. Les points forts de cette formation résident dans la qualité des enseignants et leur expertise. Les projets pratiques et les travaux de groupe encouragent l'application des connaissances acquises, ce qui renforce l'apprentissage et la compréhension des concepts

J'ai apprécié la formation, niveau prérequis je n'ai pas fait les modules obligatoires de S8 et pourtant je ne me suis pas sentie perdue quand je suis revenue en S9. Les difficultés étaient pour moi adaptées au niveau que j'avais

Les cours étaient tous très intéressants. Il y a certains cours pour lesquels la quantité de travail (en dehors du cours) était un peu trop conséquente. Je trouve que les profs sont tous très présents et toujours disponibles pour nous aider ou nous conseiller

Intéressant de voir tout ce qu'on peut faire juste avec certains logiciels

J'ai beaucoup aimé ce M2. J'ai appris beaucoup de choses. J'ai trouvé que le rythme des cours était ok même s'il y a des semaines avec plusieurs DM en parallèles. J'aime beaucoup l'ambiance de la promo et de nos relations avec l'équipe enseignante. Je me sens libre de poser mes questions. Après un an de césure j'appréhendais pas mal les premiers cours. Au final ça a été

Je ne regrette pas du tout de m'être orienté dans cette spécialisation. Ça répond parfaitement à mes attentes. Les projets thématiques et soutenances nous préparent mieux pour affronter le monde d'entreprise. Les conférences professionnelles ont été très bénéfiques. Le voyage d'études sur Vannes permettait de savoir les applications de Data Science à d'autres champs d'études, en parallèle avec nos modules du semestre 9. Bonne disponibilité et bienveillance des enseignants. Merci beaucoup de nous partager les offres de stages et CDD quasiment tous les jours, on a l'embaras du choix en France, UE

Le niveau de math est très différent entre chaque personne en fonction d'où il vient (prépa, post bac, etc), un cours de qq heures sur des rappels de math ou une liste des connaissances en math à avoir pour la rentrée pourrait pallier ce problème. De même pour Python. Le partage des offres de stage est super sympa aussi. Merci à tous pour votre disponibilité et bienveillance !

J'ai beaucoup apprécié suivre la formation. L'équipe pédagogique est très à l'écoute et toujours prête à nous aider. Néanmoins, je pense que donner une liste de recommandation de révision à faire avant la rentrée serait un plus notamment sur l'aspect statistiques. C'est possible de faire sans mais ça permettrait d'être plus à l'aise pour ceux qui souhaitent travailler un peu avant la spécialisation, notamment pour les étudiants qui ne viennent pas de l'Institut Agro Rennes ou qui reviennent de césure