

# Réunion d'information

## Présentation du M2 Data Science

### Santé, Assurance, Finance



Mercredi 04 septembre 2024



<http://www.math-evry.cnrs.fr/departement/doku.php?id=formation:master:m2ds>

L'université Paris-Saclay

Elle est née d'un regroupement :

- Université Paris-Sud
- CentraleSupélec, AgroParisTech, Ecole normale Paris-Saclay et Institut d'optique graduate school
- Université d'Evry et de Versailles-Saint-Quentin (fusion en 2025)
- Des organismes nationaux de recherche (CEA, INRIA, INSERM, INRAE, CNRS, ONERA,...)
- L'organisme coordonne l'alliance européenne EUGLOH et est membre de réseaux européens et internationaux (EUGLOH, CESAER,EUA,EUF,LERU, U7+)

C'est aussi

- Première université de France
- 220 laboratoires
- 13 000 publications par an
- 500 plateformes expérimentales
- 335 contrats avec les entreprises
- 13% du potentiel de recherche français
- 48 000 étudiants,
- 10 500 chercheurs et enseignants-chercheurs et chercheurs
- 11 000 personnels techniques et administratifs
- 8 000 étudiants et étudiantes en master
- 17 graduate school et un institut des sciences de la lumière
- 4 500 doctorant.e.s
- Classée 12<sup>ème</sup> université au classement de Shanghai
- Classée 1<sup>ère</sup> en mathématiques

# La mention de master de Mathématiques et Applications de l'Université Paris-Saclay

- 27 M1 et M2 différents couvrant tout le spectre des mathématiques fondamentales et appliquées
- 750 étudiant · e · s de Master par an
- 50 % des étudiant · e · s de Master poursuit en doctorat
- Excellente insertion professionnelle
- Un périmètre d'excellence avec 13 médailles Fields, 3 prix Abel, 4 prix Wolf

- 88.7 % de nos étudiants en master recommanderaient leur formation
- 82.5 % de nos étudiants ont apprécié le contenu de leur formation
- 90 % de nos étudiants estiment que le master leur a donné les bases nécessaires pour mener à bien leur projet professionnel
- 86 % : taux d'insertion des diplômé.e.s en 2020, 6 mois après avoir obtenu leur diplôme

# Structure du Master « Mathématiques et Applications »

Schéma des différents parcours possibles au sein du master "Mathématiques et Applications".



16

## Les métiers de la science des données

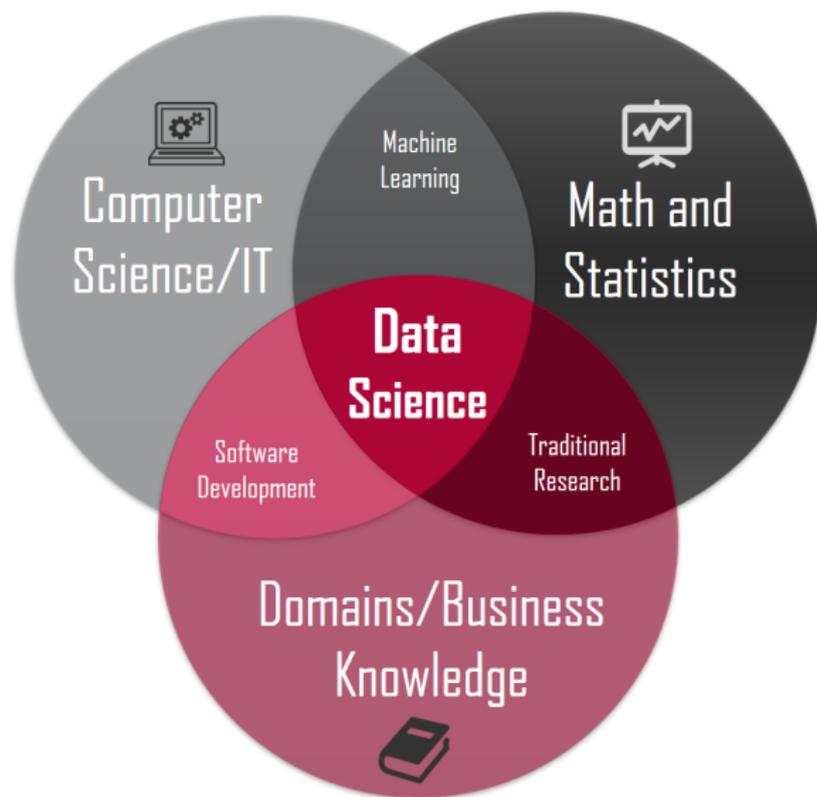


Figure 1: Source : <https://towardsdatascience.com/>

# Les métiers de la Data Science (2)



## DATA SCIENTIST

AS RARE AS UNICORNS

**Role**  
Cleans, manages and organizes (big) data

**Mindset**  
Curious data wizard

**Languages**  
R, SAS, Python, Matlab, SQL, Hive, Pig, Spark

**Skills & Talents**  
• Data Based computing  
• Predictive modeling  
• Story-telling and visualization  
• Math, Stats, Machine Learning



## STATISTICIAN

HISTORIC LEADERS OF DATA

**Role**  
Collects, analyzes and interprets qualitative as well as quantitative data with statistical theories and methods

**Mindset**  
Logical and enthusiastic stats genius

**Skills & Talents**  
• Statistical theories & methodology  
• Data mining & machine learning  
• Distributed Computing (Hadoop)

**Languages**  
R, SAS, SPSS, Matlab, Stata, Python, Perl, Hive, Pig, Spark, SQL



## DATA ANALYST

DATA DETECTIVE

**Role**  
Collects, processes and performs statistical data analyses

**Mindset**  
Innovative data junkie with high "figure-it-out" quotient

**Skills & Talents**  
• Spreadsheet tools (e.g. Excel)  
• Database systems (SQL and NO SQL based)  
• Communication & visualization  
• Math, Stats, Machine Learning

**Languages**  
R, Python, HTML, Javascript, C++, SQL



## DATA ARCHITECT

THE CONTEMPORARY DATA MODELLER

**Role**  
Creates blueprints for data management systems to integrate, centralize, protect and maintain data sources

**Mindset**  
Imagining things with a love for data architecture design patterns

**Skills & Talents**  
• Data warehousing solutions  
• In-depth knowledge of database architecture  
• ETL (Extract, Transform and Load) tools  
• Spreadsheet and BI tools  
• Data modeling  
• Systems development

**Languages**  
SQL, HAW, Hive, Pig, Spark



## DATA ENGINEER

SOFTWARE ENGINEERS BY TRADE

**Role**  
Develops, constructs, tests and maintains architectures (such as databases and large scale processing systems)

**Mindset**  
All-purpose engineer

**Skills & Talents**  
• Database systems (SQL, NO SQL, NoSQL)  
• Data modeling & ETL tools  
• Data APIs  
• Data warehousing solutions

**Languages**  
SQL, Hive, Pig, R, Matlab, SAS, SPSS, Python, Java, Ruby, C#, Perl



## BUSINESS ANALYST

CHANGE AGENT

**Role**  
Improves business process as intermediary between business and IT

**Mindset**  
Resilient project juggler

**Skills & Talents**  
• Data tools (e.g. MS Office)  
• Data visualization tools (e.g. Tableau)  
• Consistent reporting and storytelling  
• Business intelligence understanding  
• Data modeling

**Languages**  
SQL



## DATA AND ANALYTICS MANAGER

DATA SCIENCE TEAM LEADER

**Role**  
Manages a team of analysts and data scientists

**Mindset**  
Data Wizard's Cheerleader

**Skills & Talents**  
• Database systems (SQL and NO-SQL based)  
• Leadership & project management  
• Interpersonal communication  
• Data mining & predictive modeling

**Languages**  
SQL, R, SAS, Python, Matlab, Java

- Le **Data Scientist** “Un Data Scientist est un profil hybride rare : un informaticien doté des capacités de programmation nécessaires à la création de logiciels permettant de gratter, de combiner et de gérer des données provenant de diverses sources, ainsi qu'un statisticien qui sait comment tirer parti des informations qu'elles contiennent”. (Jake Porway, fondateur de DataKind)
- Le **Data Analyst** Son rôle est de comprendre les diverses données d'une entreprise (chiffres de vente, études de marché, circuits logistiques, coûts des fournisseurs, etc...) afin d'en tirer des informations claires et utiles pour l'entreprise.
- Le **Data Architect** Le Data Architect a pour rôle de construire les fondations de tous les systèmes de gestion, de prédiction, et d'analyse dont les données sont la ressource principale.
- Le **Software Engineer** Le Software Engineer a pour rôle de construire d'implémenter et de passer à l'échelle les algorithmes développés par le Data Scientist.

## ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



## MACHINE LEARNING

Machine learning begins to flourish.



## DEEP LEARNING

Deep learning breakthroughs drive AI boom.



1950's

1960's

1970's

1980's

1990's

2000's

2010's

## L'organisation du Master 2 Data Science, Santé, Assurance, Finance

- La promotion comporte 2 groupes
  - en formation initiale (FI)
  - en apprentissage (FA)
  - La différence entre les enseignements dans les deux groupes:
    - La durée du stage: 4 à 6 mois pour les FI et toute l'année, 3 jours par semaine pour les FA.
    - Le nombre d'ECTS pour le stage: 15 pour les FI et 19 pour les FA
    - Le Data Camp n'est que pour les FI
    - Les volumes horaires des cours de statistique non paramétrique et statistique asymptotique.
    - Le cours de Machine Learning de l'ENSIIE pour les étudiants qui n'ont jamais fait de machine learning.
- Contacts:
  - Marie-Luce Taupin (marieluce.taupin@univ-evry.fr): Responsable du M2 pour 2024-2025
  - Laura Djintangar (laura.djintangar@univ-evry.fr): Secrétariat pédagogique
  - Florence Varnerot (f.varnerot@cfa-eve.fr): correspondante CFA-EVE pour les étudiants en FA
- Entretiens individuels en cours d'année pour faire le point.

L'année est composée de 5 blocs (4 blocs pédagogiques: environ 500 heures+stage) :

- 1 Anglais
- 2 Statistique et machine Learning
- 3 Informatique et data camp (FI uniquement)
- 4 Options (Santé, Assurance, Finance)
- 5 Stage ou alternance

- English Semestre 1: 18h, cours à l'UEVE (pour l'option Santé et l'option Assurance), 24h (pour l'option finance)
- English Semestre 2 18h, cours à l'UEVE (pour l'option Santé et l'option Assurance), 24h (pour l'option finance)

- UE Statistics (6ECTS)
  - Nonparametric statistics (30h en FI, ML Taupin)
  - Asymptotic statistics and GLM (30h en FI, ML Taupin)
  - Statistics for stochastic processes (18h, A. Gloter)
- UE Machine learning (9 ECTS)
  - Discrete Latent Variables (18h, N. Jouvin)
  - Deep learning (30h, B. Hanczar et K. Dradjat)
  - [Suivant votre parcours précédent](#) :
    - Machine learning 1 (21h, cours à l'ENSIIE, M. Mougeot) **OU**
    - Visualisation and dimension reduction (18h, C. Ambroise)
- UE Advanced statistics and machine learning (6ECTS)
  - Optimization for Data Science (21h, M. Castella)
  - Bayesian Stat (18h, G. Kon-Kam-King)
  - Graphical models (18h, C. Ambroise)

- UE Basics of Informatics (2 ECTS)
  - Introduction to R (12h, M.L. Taupin)
  - Introduction to Databases (12h, F. Samson)
- UE Advanced Informatics (4 ECTS)+4 ECTS (FI uniquement)
  - Algorithmique et GPU (27h, V. Runge et A. Gloter)
  - Big Data / Advanced topics for databases(18h, S. Dumbrava)
  - Data Camp (9h, N. Jouvin, **FI uniquement**)

- UE Option S1 (5 ECTS)
  - CHOIX 1 pour l'option santé
    - Introduction to genetics (18h, E Teixeira et V. Chaudru)
    - Longitudinal data analysis (18h, J. Park)
  - CHOIX 2 pour l'option assurance
    - Financial Econometry (15h, J. Park )
    - Longitudinal data analysis (21h, J. Park)
  - CHOIX 3 pour l'option finance
    - Finance numérique (42h, A. Kebaier)
- UE Option S2 (5 ECTS)
  - CHOIX 1 pour l'option santé
    - Statistiques pour la génétique et la génomique (18h, C. Dalmasso)
    - Bioinformatics (18h, M.L. Martin et V. Brunaud)
  - CHOIX 2 pour les options finance et assurance
    - Machine learning pour la finance et l'assurance (36h, A. Kebaier)

- Pour les FI (15 ECTS)
  - Durée de 16 à 24 semaines à partir du mois d'avril.
  - La thématique doit être validée par un responsable du master et un responsable de votre école pour les double cursus
  - Une convention de stage est obligatoire entre les différentes parties (contact Laura Djintangar [laura.djintangar@univ-evry.fr](mailto:laura.djintangar@univ-evry.fr))
- Pour les FA (19 ECTS)
  - Alternance à l'année les lundis et mardis (et les lundis, mardis et mercredis à partir de mars)
  - La thématique doit être validée par un des responsables
  - Un tuteur universitaire sera désigné.
  - Une convention de stage est obligatoire entre les différentes parties (contact Laura Djintangar [laura.djintangar@univ-evry.fr](mailto:laura.djintangar@univ-evry.fr))

- L'organisation de l'alternance
  - Lundi et mardi en entreprise jusqu'en mars
  - Les lundis, mardis et mercredis en entreprise à partir de mars jusque mi-avril
  - Mercredi, jeudi et vendredi à l'université jusqu'en mars (cours, TP, préparation de projets, examens, ...)
  - Jeudi et vendredi à l'université à partir de mars jusque mi-avril.
  - Vous pouvez consulter l'EDT avec les salles à l'adresse <https://edt-univ-evry.hyperplanning.fr/hp/?login=true> (espace invité)
  - Le M2 a un site web: <http://www.math-evry.cnrs.fr/departement/doku.php?id=formation:master:m2ds>
  - Sur le site web, vous trouverez des documents ainsi qu'un agenda google.

## Règlement des études

- les UEs d'un même bloc sont compensables si la note minimale (seuil de compensation) de 7/20 pour chaque UE a été atteinte
- les blocs sont compensables et validés si moyenne des UEs du bloc est supérieure ou égale à 10.
- à l'exception du **stage qui n'est ni compensable ni compensant**

## Exemples

- Pas de compensation dans le bloc ni avec un autre bloc

Intitulé de l'UE	Note	Bloc	
English S4	6	1: English	→ rattrapage
English S3	14	1: English	
Moyenne	10		

- Compensation possible avec un autre bloc

Intitulé de l'UE	Note	Bloc	
English S4	7	1: English	→ rattrapage si pas de compensation avec d'autres blocs
English S3	11	1: English	
Moyenne	9		

- Rapport : doit faire entre 20 et 40 pages, si possible rédigé en Latex.  
Pour chaque rapport, les points importants sont :
- L'exposé clair d'une problématique scientifique (ou de plusieurs) et des méthodes déployées pour y répondre
- Une discussion de la pertinence des résultats obtenus
- L'utilisation de références bibliographiques
- L'orthographe, la syntaxe (oui c'est important)
- Dans le rapport, il est très important de citer vos sources ( section bibliographie ), de mettre un lien vers votre code (GitHub ou en annexe) et de détailler certains calculs techniques en annexe si besoin.
- Des articles ou des livres doivent être cités, pas seulement des sources internet.
- De préférence, le rapport doit être rédigé en latex.
- Organisation typique du rapport
  - 1 Introduction avec présentation de l'entreprise et de l'équipe (3 pages maximum) , puis présentation de la problématique scientifique
  - 2 Présentation des méthodes utilisées pour répondre à la (ou les) question(s)/problématique(s)
  - 3 Il faut aussi présenter les méthodes qui n'ont pas été concluantes mais importantes dans la démarche)
  - 4 Résultats obtenus
  - 5 Conclusion et perspectives

(Attention, le plagiat est sévèrement puni, un paragraphe recopié est facilement détectable)

- Soutenance (45 min par étudiant)
  - 30 min d'exposé
  - 10 min de questions
  - 5 min de discussion avec l'encadrant entreprise
  - La soutenance doit être en présentiel pour l'étudiant
  - Distanciel possible pour l'encadrant.e
  - La soutenance reprend un ou des points particulièrement pertinents du rapport.

Après le Master M2 DS

Trois possibilités principales:

- Emploi en entreprise;
- Doctorat de type thèse Cifre (à moitié en entreprise et à moitié à l'université): Cette option nécessite de trouver une entreprise, un.e encadrant.e académique et que les deux se mettent d'accord sur un sujet de thèse qui correspond à des besoins de l'entreprise et qui soit aussi susceptible de proposer des développements méthodologiques/théoriques suffisants pour une thèse. Le processus peut-être assez long et il vaut donc mieux s'y prendre à l'avance pour un début en septembre 2025 ou janvier 2026;
- Doctorat de type thèse académique: Il faut se mettre rapidement à la recherche d'un.e encadrant.e et d'un sujet. Vous pouvez par exemple contacter des collègues enseignants-chercheurs du laboratoire ou d'autres laboratoires. Comme pour les thèses Cifre, il faut s'y prendre relativement à l'avance.
- Attention: il n'est pas possible de faire une thèse sans financement. Les financements de thèse du ministère, de la région, des écoles doctorales, etc... se décident vers avril-mai 2025 pour un début de thèse à l'automne 2025.
- Si votre projet est de faire un doctorat à l'issue du M2, le mieux est de me contacter le plus rapidement possible en me précisant où en est votre recherche.

Comment me joindre?

- Mon bureau est au 4 ème étage de l'IBGBI (bureau 425) mais vous ne pouvez en général pas y accéder sans RDV (badge à l'entrée du couloir)
- Le mieux est de m'envoyer un mail pour que nous convenions d'un RDV.

Avez-vous des questions?

Bonne rentrée!