

INFORMATIONS PRATIQUES

LABORATOIRES

- CMAP : CNRS & Ecole polytechnique
- LTCI : CNRS & Télécom Paris Tech
- LMO : CNRS & UPSud
- CMLA : CNRS & ENS Cachan
- CREST : CNRS, INSEE & ENSAE
- LRI : CNRS & UPSud
- LIX : CNRS & Ecole polytechnique

PARTENAIRES SOCIO-ÉCONOMIQUES

- Orange, Thales, Keyrus, Deloitte, Yves Rocher, voyages-sncf.com, Criteo, PSA Peugeot Citroën, Safran

ETABLISSEMENTS PARTENAIRES



- Ecole Polytechnique, Route de Saclay, 91128 Palaiseau.



- Télécom ParisTech, 46 Rue Barrault, 75013 Paris.



- Université Paris Sud, 15 Rue Georges Clémenceau, 91405 Orsay.

Comprendre le monde, construire l'avenir



- ENS Cachan, 61 Avenue du Président Wilson, 94230 Cachan.



- ENSAE, 3 Avenue Pierre Larousse, 92240 Malakoff.

CONTACTS

Secrétariat

- Sandra Schnakenbourg, sandra@cmap.polytechnique.fr

Responsables

- Eric Moulines, eric.moulines@polytechnique.edu
- Erwan Le Pennec, erwan.Le-Pennec@polytechnique.edu
- Florence d'Alché-Buc, florence.dalche@telecom-paristech.fr



www.universite-paris-saclay.fr

Conception graphique : Université Paris-Sud - Janvier 2016

université
PARIS-SACLAY

SCHOOL

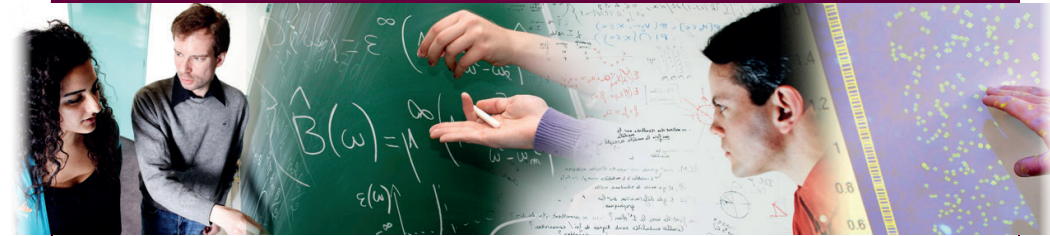
SCIENCES
FONDAMENTALES

MASTER

Mathématiques
et applications

Master 2 Mathématiques et applications

PARCOURS : Data Sciences



Les data sciences, qui mélangent modélisation mathématique, statistique, informatique, visualisation et applications ont pour objectif de passer du stockage et de la diffusion de l'information à la création de connaissances. Ce passage des données aux connaissances est porteur de nombreux défis qui requièrent une approche interdisciplinaire. Les data sciences s'appuient fortement sur le traitement statistique de l'information (statistiques mathématiques, statistiques numériques, apprentissage statistique ou machine learning). De l'analyse de données exploratoires aux techniques les plus sophistiquées d'inférence (modèles graphiques hiérarchiques, deep learning, machine à vecteurs de support), l'ensemble des méthodes statistiques des plus éprouvées aux plus modernes sont exploitées. Le « Big Data » marque le début d'une transformation majeure, qui va affecter de façon profonde l'ensemble des secteurs (de l'e-commerce à la recherche scientifique en passant par la finance et la santé !).



PRÉ-REQUIS

Des compétences, en probabilité, en statistique et en informatique sont recommandées.

OBJECTIFS PÉDAGOGIQUES

Les méthodes d'extraction de connaissances, pour pouvoir être développées à l'échelle de masses de données requièrent la maîtrise des mécanismes de parallélisation et distribution des calculs, des méthodes d'accès et de requêtes à des bases de données réparties à très grande échelle et en temps-réel. La grande dimension influence la conception même des algorithmes d'extraction de connaissance et d'inférence statistique, poussant à l'utilisation de nouveaux outils issus de différentes branches des mathématiques (analyse fonctionnelle, analyse numérique, optimisation convexe et non convexe) dont il s'agit d'acquérir la compréhension.

Ce parcours associe des cours théoriques et méthodologiques complétés par des projets en « vrai grandeur » faisant intervenir tous les aspects des sciences des données, depuis l'acquisition jusqu'à l'exploitation et l'analyse. Une partie significative du parcours sera validée sous forme de projets. Une des originalités de ce parcours: un recours à des pédagogies innovantes basés sur l'apprentissage par projets et la participation à des compétitions de data sciences (kaggle).

Le candidat choisit librement un stage proposé par l'un des enseignants du master, un stage en entreprise proposé dans le cadre de la « bourse des stages », ou un stage d'origine différente ayant reçu l'agrément d'un enseignant du master. Le stage doit être effectué après l'inscription au master. Il doit présenter un enjeu scientifique réel et le développement applicatif d'un des thèmes développés dans le master. La durée est de quatre mois minimum, et débute au 1er avril.

DÉBOUCHÉS

Il existe actuellement un large déficit d'ingénieurs de très « haut-niveau » en datasciences.

Les compétences dans le domaine de l'apprentissage statistique et du traitement de données massives (Big Data) que vont acquérir les étudiants suivant ce parcours sont recherchées aussi bien dans des start-up (dont beaucoup ont des projets basés sur l'extraction de connaissance, les méthodes de recommandation, de ciblage) que dans des grandes entreprises (tous les domaines d'activité sont impactés). Ces nouveaux métiers de « datascientists » sont multiformes, ils vont de la mise en place de nouvelles générations de systèmes d'informations décisionnels aux développements d'applications complètement nouvelles (autour du e-commerce, de la recommandation, du minage de réseaux sociaux, etc..).

Le besoin de doctorants est également important dans ce domaine d'innovations de rupture. Les propositions de thèses sont nombreuses dans la recherche publique (Université, CNRS, INRIA, CEA, CNES, INRA, INSERM, LETI, etc.) et dans les grands laboratoires de recherche dans l'industrie (Aérospatiale, Alcatel, Sagem, General Electric, Matra, Philips, Siemens, Thales, EDF, etc.).

RECHERCHE

La formation est adossée aux laboratoires de mathématiques ou d'informatique des établissements partenaires.