

Le master Data Science : Santé, assurance, finance



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

Big data / data science : un tour d'horizon

Un nouveau contexte

Des données partout

- ▶ Volume gigantesque,
- ▶ Très grande variété...

Unités de calcul abordables

- ▶ Cloud computing
 - ▶ Graphical Processor Units (GPU)...
-
- ▶ Intérêt académique et industriel croissant !

Data never sleeps

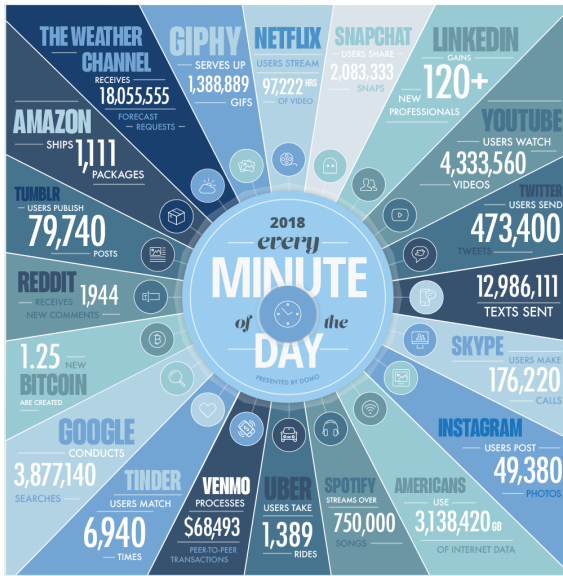


Figure 1: <https://www.domo.com/>

Les sources de données

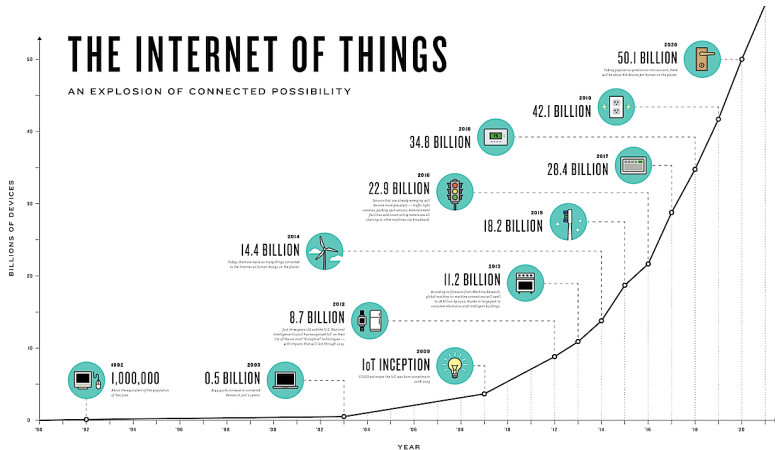


Figure 2: <http://www.ncta.com/>

3Vs du Big Data

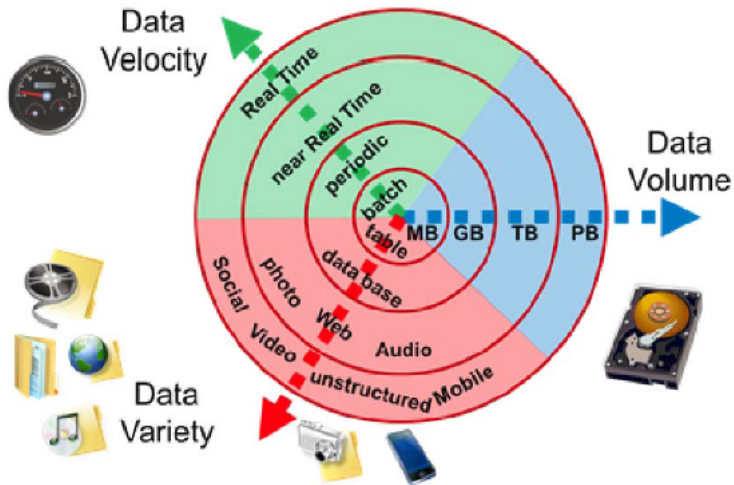


Figure 3: Source : wikipedia

Les métiers de la Data Science (1)

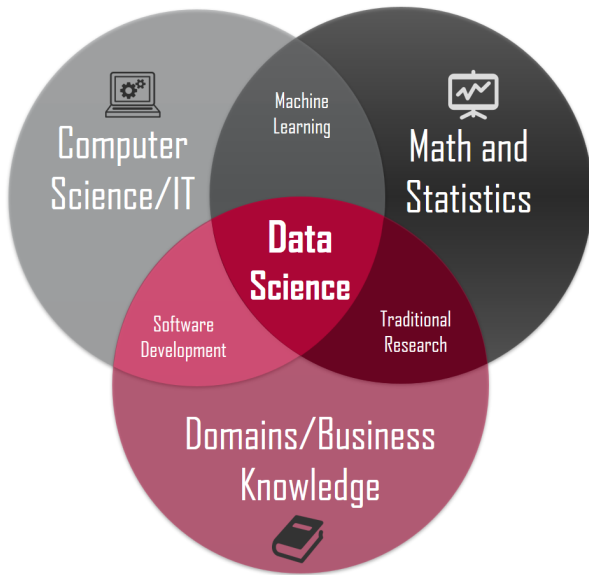


Figure 4: Source : <https://towardsdatascience.com/>

Les métiers de la Data Science (2)

- ▶ 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) Voir par exemple lien linkedin
- ▶ 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.
Voir par exemple lien linkedin
- ▶ 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. Voir par exemple lien linkedin
- ▶ 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. lien linkedin

Les métiers de la Data Science (3)



Figure 5: Source : <https://www.houseofbots.com/>

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

M2 Data Science-site EVRY, Semestre 1 (30 ECTS) I

UE 1: Fundamentals (4 ECTS)

- English (18h TD, 2ECTS, cours à l'UEVE)

UE 2: Statistics (8 ECTS)

- Suivant votre parcours précédent :
 - ▶ Introduction au machine learning (12 cours + projet, 2ECTS, cours à l'UEVE)
intervenant : A. Guilloux
 - ▶ Theoretical guidelines for high-dimensional data analysis (20h cours, 2 ECTS cours à Saclay)
intervenant : C. Giraud
- Longitudinal data analysis (21h, 2 ECTS, cours à l'ENSIIE)
intervenant : A. Guilloux
- Statistics for stochastic processes (18h cours, 2ECTS, cours à l'UEVE)
intervenant : A. Gloter
- Nonparametric statistics (18h, 2ECTS, cours à l'UEVE)
intervenant : ML Taupin

M2 Data Science-site EVRY, Semestre 1 (30 ECTS)

UE 3 Informatics (6 ECTS)

- Introduction to Databases (9h cours/TD, 2 ECTS, cours à l'UEVE)
intervenant : M. Szafranski
- Scientific Python, R, Calcul parallèle GPU (15h, 2 ECTS cours à l'UEVE)
intervenants : V. Runge, M. Pascucci, A. Gloter
- Advanced topics in Databases (18h, 2 ECTS, cours à l'UEVE)
intervenant : Y. Sebiat

UE 4 Machine Learning I (8 ECTS)

- Unsupervised learning (18h, 2 ECTS, cours à l'UEVE)
intervenant : C. Ambroise
- Machine learning (42h, 4 ECTS, cours à l'ENSIIE)
intervenants : M. Mougeot, C. Ambroise
- Optimization for Data Science (21h, 2 ECTS, cours à TSP)
intervenants : M. Castella

UE 5: Options I (6 ECTS à valider)

► CHOIX 1 : santé

- Statistics for genetics and genomics (25h, 2 ECTS, cours à l'UEVE)
intervenant : à définir, cours électif
- Genetics et analyse de liaison (21h, 2 ECTS, cours à l'UEVE)
intervenant : E Teixeira et V. Chaudru
- Méthodologie d'analyse des données d'expression (25h, 2 ECTS, cours à l'UEVE)
intervenant : à définir, cours électif

► CHOIX 2 : assurance

- Gestion des risques (36h cours+ 15h Td, 4 ECTS, cours à l'UEVE)
intervenant : V. Roncalli, cours électif

► CHOIX 3 : finance

- Financial Econometry (21cours/TD, 2 ECTS)
intervenant : A. Gloter, cours électif
- Méthodes numériques de pricing (42h cours, 4 ECTS, cours à l'UEVE)
intervenant : S. Crépey, cours électif

UE 6: Fundamentals (2ECTS)

- English (18h TD, 2ECTS, cours à l'UEVE)

UE 7 Advanced Statistics and Machine Learning (6 ECTS)

- Deep learning (10h + projet, 2ECTS, cours à l'UEVE)
intervenants : B. Hanczar
- Computational statistics (21h cours/TD, 2 ECTS, cours à TSP)
intervenants : R. Douc, Y. Petetin, F. Debouvries
- Graphical models (9hCM+9hTD, 2 ECTS, cours à l'UEVE)
intervenants : Christophe Ambroise

UE 8 Options II (6ECTS)

► CHOIX 1 : santé

- Bayesian methods (18h cours/TD, 2 ECTS, cours à l'UEVE)
intervenants : E. Kuhn
- Multiple testing and resampling (18h cours/TD, 2,5 ECTS, cours à l'UEVE)
intervenants : C. Dalmasso et A. Guilloux
- Bioinformatics (18h, 2ECTS, cours à l'UEVE)
intervenants : C. Rizzon et Y. Diaz

► CHOIX 2 : assurance

- Assurance (54h cours, 6 ECTS, cours à l'UEVE)
intervenants : T. Roncalli, cours électif

► CHOIX 3 : finance

- IT quant (51h cours, 6 ECTS, cours à l'UEVE)
intervenants : S. Crépey, cours électif

UE 9 Data camp (4 ECTS) (20h) intervenants : S. Bussy et A. Guilloux

UE 10 Stage (12 ECTS)

- ▶ Durée : minimum 16 à 24 semaines à partir du mois d'avril
- ▶ La thématique doit être validée par Agathe Guilloux et un responsable de votre école pour les double cursus
- ▶ Une convention de stage est obligatoire entre les différentes parties (contact Adelia Soarez adelia.soarez@univ-evry.fr)

Règlement des études de l'Université Paris-Saclay

Une année de formation compte au moins 60 ECTS. Les conditions sont cumulatives :

- ▶ les semestres ne sont pas compensables entre eux
- ▶ la note minimale pour la compensation est de 7/20 pour chaque UE
- ▶ toutes les UE du premier semestre sont compensables entre elles
- ▶ c'est aussi le cas des UE du deuxième semestre
- ▶ à l'exception du **stage qui n'est ni compensable ni compensant**

Plan du campus d'Evry



Démarches administratives

Votre inscription est obligatoire :

- ▶ Inscription administrative : à l'université d'Evry val d'Essonne ou auprès de votre établissement d'origine pour les étudiants de l'ENSIIE, Télécom Sud Paris)
- ▶ Inscription pédagogique obligatoire sur l'outil notes de Paris Saclay (mail envoyé début octobre)

Prêt/achat d'ordinateur

L'université d'Evry lance sa campagne de financement pour l'achat d'un ordinateur portable et sa campagne de prêt d'ordinateur portable.

- ▶ Aide à l'achat d'un ordinateur portable (critères de sélection : Niveau des ressources, Lettre de motivation, Résultats pédagogiques de l'année antérieure)
- ▶ Prêt d'un ordinateur portable (critères de sélection : Lettre de motivation (projet d'études et professionnel, métier envisagé), résultats pédagogiques de l'année antérieure)

Dossier à remplir en ligne sur le site de l'Université :

<https://www.univ-evry.fr/vie-de-campus/vie-pratique/bourses-et-aides.html>

Ouverts aux étudiants de l'Université d'Evry.

Contact : direction de la Vie Étudiante vie.etudiante@univ-evry.fr

Tout dossier incomplet ne sera pas examiné. Le dossier est à remplir au **plus tard pour le vendredi 25 octobre 2019**.

Configuration de votre ordinateur

Il faut installer sur votre ordinateur

- ▶ R : version > 3.3 cf : <https://cran.r-project.org>
- ▶ Rstudio : version > 1 cf : <https://www.rstudio.com/products/RStudio/>
- ▶ Python : anaconda distribution (> 4) avec python 3
<https://docs.continuum.io/anaconda/install/>

Si vous voulez approfondir l'un ou l'autre

- ▶ R for Data Science de Garrett Grolemund et Hadley Wickham
<http://r4ds.had.co.nz>
- ▶ Python Data Science handbook de Jake VanderPlas
<https://github.com/jakevdp/PythonDataScienceHandbook>

Autres démarches

- ▶ **Chaîne slack**

https://join.slack.com/t/m2dssaf/shared_invite/enQtNzM3MTkxNjQzMTI2LTM

- ▶ **Choix des options** via la questionnaire en ligne :

https://docs.google.com/forms/d/e/1FAIpQLSfxRWi8EbpUmAginSO-Q9mZjJzRRdsf5KpuEEFYswZfGHZAHQ/viewform?usp=sf_link

- ▶ **Google colab** : dans certaines matières sera utilisé

<https://colab.research.google.com> qui nécessite un compte gmail.

Pour demain

- ▶ Vous aurez besoin de python via anaconda ou Google colab
- ▶ Vous devez récupérer les notebooks et les données “gro”
- ▶ Vous devez faire seul (et comprendre ...) les notebooks
"intro_python.ipynb" et "intro_numpy.ipynb"
- ▶ Vous devez commencer le notebook "ds_with_python_1.ipynb" jusqu'à
l'import des données (5ième cellule de code)