

Présentation du monde de la data & de DataSciEence

Première formation de l'année scolaire 2022/2023

Association DataSciEence

Formations en Data Science & Artificial Intelligence

Jeudi 6 Octobre 2022 - 16h00 à 17h30

datascience@gmail.com • datascience-ensiie.github.io • datascience.iiens.net



Présentation de la séance

La présentation sera découpée selon le plan suivant :

Monde de la Data Science

Association DataSciEnce

Data Science à l'ENSIIE

SI de l'association

Conclusion

Présentation du monde la data

→ **Qu'est ce que la Data Science ?**

- ▶ La **Data Science** recouvre tous les aspects du **traitement** des données, de la **collecte**, en passant par la **modélisation**, jusqu'aux **prédictions** ;
- ▶ Elle permet d'explorer et d'analyser des grandes quantités de données (ou pas) pour en tirer des informations utiles ;
- ▶ C'est une discipline qui mêle principalement **statistiques**, **informatique** et **connaissances du thème abordé** (finance, santé, aéronautique...).

Figure: Les trois piliers de la Data Science

Présentation du monde la data

→ **Qu'est ce que la Data Science ?**

- ▶ La **Data Science** recouvre tous les aspects du **traitement** des données, de la **collecte**, en passant par la **modélisation**, jusqu'aux **prédictions** ;
- ▶ Elle permet d'explorer et d'analyser des grandes quantités de données (ou pas) pour en tirer des informations utiles ;
- ▶ C'est une discipline qui mêle principalement **statistiques**, **informatique** et **connaissances du thème abordé** (finance, santé, aéronautique...).

Figure: Les trois piliers de la Data Science

Présentation du monde la data

→ **Qu'est ce que la Data Science ?**

- ▶ La **Data Science** recouvre tous les aspects du **traitement** des données, de la **collecte**, en passant par la **modélisation**, jusqu'aux **prédictions** ;
- ▶ Elle permet d'explorer et d'analyser des grandes quantités de données (ou pas) pour en tirer des informations utiles ;
- ▶ C'est une discipline qui mêle principalement **statistiques**, **informatique** et **connaissances du thème abordé** (finance, santé, aéronautique...).

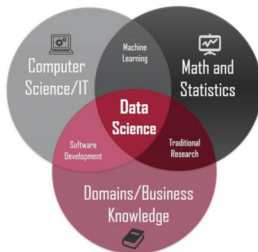


Figure: Les trois piliers de la Data Science

Présentation du monde la data

→ Les différents aspects de la Data Science

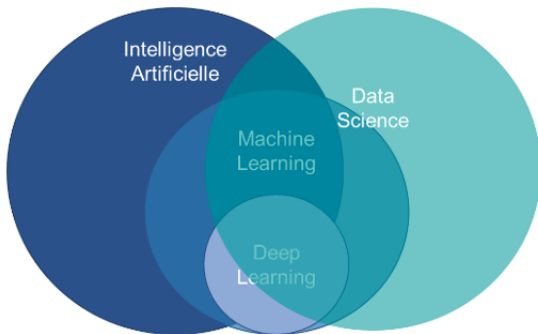


Figure: Les différents domaines de la Data

- ▶ **IA:** les algorithmes vont simuler la pensée et le raisonnement d'un humain
- ▶ **ML:** les algorithmes vont chercher à prédire ou relier des données entre elles
- ▶ **DL:** basé sur les réseaux de neurones, nécessite moins d'intervention humaine

Présentation du monde la data

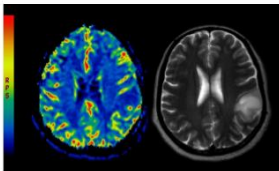
→ Exemple de différentes applications de la Data Science



Analyse des dangers



Prédire les cours de marché



Prévention des tumeurs



Prédire les prix de l'immobilier

Présentation du monde la data

→ **Pourquoi ce domaine est-il en plein essor ?**

- ▶ **Quantité considérable de données au sein des entreprises**
- ▶ Exploitation de ces données par les entreprises est un véritable enjeu
- ▶ Les applications de la Data Science répondent à cette problématique
- ▶ Accès à plein d'informations "cachées" dans les données brutes

Présentation du monde la data

→ **Pourquoi ce domaine est-il en plein essor ?**

- ▶ Quantité considérable de données au sein des entreprises
- ▶ Exploitation de ces données par les entreprises est un véritable enjeu
- ▶ Les applications de la Data Science répondent à cette problématique
- ▶ Accès à plein d'informations "cachées" dans les données brutes

Présentation du monde la data

→ **Pourquoi ce domaine est-il en plein essor ?**

- ▶ Quantité considérable de données au sein des entreprises
- ▶ Exploitation de ces données par les entreprises est un véritable enjeu
- ▶ Les applications de la Data Science répondent à cette problématique
- ▶ Accès à plein d'informations "cachées" dans les données brutes

Présentation du monde la data

→ **Pourquoi ce domaine est-il en plein essor ?**

- ▶ Quantité considérable de données au sein des entreprises
- ▶ Exploitation de ces données par les entreprises est un véritable enjeu
- ▶ Les applications de la Data Science répondent à cette problématique
- ▶ Accès à plein d'informations "cachées" dans les données brutes

Présentation du monde la data

→ **Pourquoi ce domaine est-il en plein essor ?**

- ▶ Quantité considérable de données au sein des entreprises
- ▶ Exploitation de ces données par les entreprises est un véritable enjeu
- ▶ Les applications de la Data Science répondent à cette problématique
- ▶ Accès à plein d'informations "cachées" dans les données brutes

Devenu très à la mode aujourd'hui...

Présentation du monde la data

→ Différences entre Data Analyst, Data Scientist et Data Engineer

BUT = DONNER DE LA VALEUR AUX DONNÉES

- ▶ **Data Analyst** : transforme les données brutes en informations exploitables (*extraction, nettoyage, visualisation*)
- ▶ **Data Scientist** : le plus polyvalent, s'occupe des modélisations (*mathématiques et informatiques*) et de l'analyses des prédictions
- ▶ **Data Engineer** : développe, met en place et maintient les outils qui servent aux analyses (*base de données, gestion de version...*)
- ▶ **Machine Learning Engineer** : proche du Data Scientist, conçoit des logiciels permettant d'automatiser les modèles prédictifs des analyses

Présentation du monde la data

→ Différences entre Data Analyst, Data Scientist et Data Engineer

BUT = DONNER DE LA VALEUR AUX DONNÉES

- ▶ **Data Analyst** : transforme les données brutes en informations exploitables (*extraction, nettoyage, visualisation*)
- ▶ **Data Scientist** : le plus polyvalent, s'occupe des modélisations (*mathématiques et informatiques*) et de l'analyses des prédictions
- ▶ **Data Engineer** : développe, met en place et maintient les outils qui servent aux analyses (*base de données, gestion de version...*)
- ▶ **Machine Learning Engineer** : proche du Data Scientist, conçoit des logiciels permettant d'automatiser les modèles prédictifs des analyses

Présentation du monde la data

→ Différences entre Data Analyst, Data Scientist et Data Engineer

BUT = DONNER DE LA VALEUR AUX DONNÉES

- ▶ **Data Analyst** : transforme les données brutes en informations exploitables (*extraction, nettoyage, visualisation*)
- ▶ **Data Scientist** : le plus polyvalent, s'occupe des modélisations (*mathématiques et informatiques*) et de l'analyses des prédictions
- ▶ **Data Engineer** : développe, met en place et maintient les outils qui servent aux analyses (*base de données, gestion de version...*)
- ▶ **Machine Learning Engineer** : proche du Data Scientist, conçoit des logiciels permettant d'automatiser les modèles prédictifs des analyses

Présentation du monde la data

→ Différences entre Data Analyst, Data Scientist et Data Engineer

BUT = DONNER DE LA VALEUR AUX DONNÉES

- ▶ **Data Analyst** : transforme les données brutes en informations exploitables (*extraction, nettoyage, visualisation*)
- ▶ **Data Scientist** : le plus polyvalent, s'occupe des modélisations (*mathématiques et informatiques*) et de l'analyses des prédictions
- ▶ **Data Engineer** : développe, met en place et maintient les outils qui servent aux analyses (*base de données, gestion de version...*)
- ▶ **Machine Learning Engineer** : proche du Data Scientist, conçoit des logiciels permettant d'automatiser les modèles prédictifs des analyses

Présentation de l'association

→ Présentation du mandat 2022/2023

- ▶ Association du BdE fondée en 2019 par Nicolas “Cosmos” Makaroff, alors 2A en Mathématiques Appliquées et Président de la Junior-Entreprise Dièse

Composition du bureau pour l'année scolaire 2022/2023

- ▶ Président : **Adib HABBOU**
- ▶ Vice-Présidente : **Rabab KHATIB**
- ▶ Secrétaire général : **Alae KHIDOUR**
- ▶ Trésorier: **Valentin GARDEL**
- ▶ Responsable Communication : **Rokaya BENNAZI**
- ▶ Responsable Évènements : **Hugo PICARD**

Présentation de l'association

→ **Présentation des réseaux sociaux**



Présentation de l'association

→ Objectifs de l'association

DataScIIence s'oriente autour de 4 objectifs :

Réaliser des
formations

Inviter des
intervenants

Participer à des
forums

Réaliser des projets

Présentation des formations

→ Objectif principal de l'association

Formations interactives mêlant côté théorique et côté pratique en Python

- ▶ Découverte des outils et références
- ▶ Manipulation de Data Frame (Pandas, Numpy)
- ▶ Data Visualisation (Matplotlib, Seaborn, Plotly)
- ▶ Web Scraping (BeautifulSoup, Requests, Selenium)
- ▶ Introduction au Machine Learning (Scikit-Learn)
- ▶ Natural Language Processing (NLTK, Gensim)
- ▶ Computer Vision (OpenCV, Yolo, OpenPose)
- ▶ Neural Networks (Keras, TensorFlow)
- ▶ ...

Présentation des formations DataSciEnce

→ Outils de formation

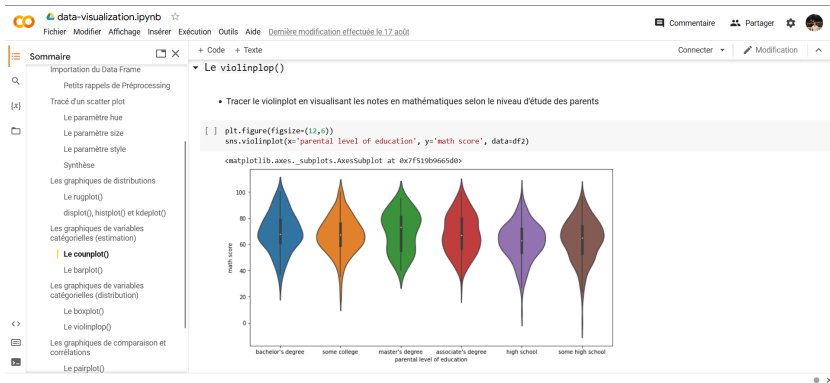


Figure: Exemple de Jupyter Notebook sur Google Colab

Projets en DataScIIEnce



Présentation des formations à l'ENSIIE

Unité d'enseignement	Responsable
(PROB) Probabilités	LY VATH
(ANNU) Méthodes d'analyse numérique	TORRI
(GROP) Graphes & Optimisation	MERABET
(CBDR) Conception de bases de données relationnelles	SZAFRANSKI
(STAT) Statistiques	BRUNEL
(OPMA) Optimisation	FAYE
(PRIM) Projet Mathématiques	PULIDO

Table: UE du parcours Mathématiques Appliquées (MA) - Semestre 1/Semestre 2

Unité d'enseignement	Responsable
(MERR) Modèles de Régression Régularisée	MOUGEOT
(PRST) Processus Stochastiques	SAGNA
(PIMA) Méthode Agile ou (ANAF) Analyse Fonctionnelle	GAUTIER / MATOS
(PRAP) Programmation Avancée & Projet	TORRI
(ANDO) Analyse de Données	AMBROISE
(REOP) Recherche Opérationnelle	WATEL

Table: UE du parcours Mathématiques Appliquées (MA) - Semestre 3

Présentation des formations à l'ENSIIE

Unité d'enseignement	Responsable
(MESI) Méthodes de Simulation	SAGNA
(MOST) Modélisation Statistique	SZAFRANSKI
(INMF) Modèles Financiers ou (CORO) Recherche Op.	LIM / FAYE
(ANEDP) Analyse des EDP ou (IPBD) Big Data	TORRI / LEBRETON
(CAST) Calcul Stochastique	BENEZET / PULDIO
(PRRE) Projet Recherche ou (PRB) Pattern Recognition	PULIDO / GARCIA

Table: UE du parcours Mathématiques Appliquées (MA) - Semestre 4

Unité d'enseignement	Responsable
(MOSC) Modél. Stoch. & Contrôle ou (PyDS) Python for DS	LIM / KOUAMO
(MALE) Machine Learning	MOUGEOT
(MENF) Méthodes numériques en Finance	KEBAIER
(INRF) Instruments Financiers	LY VATH
(MOSA) Modélisation Statistique Avancée	CHARANTONIS

Table: UE du parcours Mathématiques Appliquées (MA) - Semestre 5

ATTENTION : Certaines matières sont obligatoires pour avoir le M1 MA

Masters & Destinations possibles

Masters possibles

Nombreuses possibilités avec l'**Université Paris-Saclay** partenaire de l'**ENSIIE** :

- ▶ Finance : M2QF et M2 GRA
- ▶ Data Science : M2 DS et M2 IMSD
- ▶ Intelligence Artificielle : M2 IA
- ▶ Recherche Opérationnelle : M2 MPRO

D'autres masters sont possibles en dehors des partenariats de l'école mais ils sont plus sélectifs : MVA de l'ENS-Saclay, Master Data Science de l'X...

Destinations possibles

Partir dans une université étrangère partenaire de l'**ENSIIE** :

- ▶ **Bologne, Munich ou Florence** en Finance
- ▶ **Politecnico di Milano** en Maths Appliquées
- ▶ Expériences d'échange en Europe mais aussi en Asie ou USA/Canada

Présentation des SI

- ▶ **Discord** : Annonce des différentes formations, forums, évènements...
- ▶ **Site Internet** : Centralisation des ressources : slides, notebooks...
- ▶ **GitHub** : Centralisation du code source des formations, des slides...

Questions/Réponses

Merci pour votre attention !



Slides de la séance



Formulaire d'évaluation de la séance