



Conseil en data stratégie

**Offre de Formation
au Big Data
au Machine Learning
et à l'Intelligence Artificielle**



Publics visés

- ✓ La valorisation pérenne des données passe par un investissement dans l'humain :

1

Sensibiliser vos collaborateurs aux changements induits par le big data et le machine learning

2

Former certains d'entre eux à une compréhension voire une maîtrise approfondie de ces sujets

- ✓ À cet effet, Mémorandum propose un parcours de sensibilisation / formation adapté à 4 types de population :

	Qui ?	Quoi ?
1	Collaborateurs éloignés des questions data	Comprendre en quoi le partage des données au sein de l'entreprise peut aider tout à chacun dans son travail quotidien
2	Managers business	Devenir le sponsor et pilote de projets data opérationnels : comment choisir un projet et gérer les équipes
3	Utilisateurs de la data (statisticiens, webanalysts, responsables de reporting...)	Actualiser des compétences analytiques en expertise machine learning (gestion de carrière interne Vs recrutement)
4	Datascientists junior	Dépasser la technique : Organiser le dialogue constructif avec les experts métiers : faire émerger un besoin, partager un résultat.

Notre philosophie : être centré métier



L'enjeu principal est de former les métiers à la donnée et former les experts donnés aux métiers

- ✓ **Le plus difficile** dans la démarche data ce n'est pas de trouver la réponse (la data science est très outillée !), mais de **poser la bonne question**. Elle ne se trouve pas du premier coup et est l'affaire de tous.
- ✓ La data science
 - ✓ Peut s'appréhender de manière très concrète et profonde sans mathématique (cette compréhension est nécessaire pour doper l'interface métier data)
 - ✓ Est techniquement accessible à un large public : le code n'est plus un passage obligé
 - ✓ N'a pas besoin de « big data » pour commencer à s'exprimer
- ✓ Les projets de valorisation de données n'aboutissent pas tous, loin s'en faut. L'enjeu est de savoir détecter et décider rapidement quand le potentiel n'est pas suffisant.

Notre conviction est que les entreprises doivent former plus de pilotes du business que de mécaniciens de la donnée

Notre ambition : transmettre une culture



Neuf convictions qui sous-tendent notre philosophie

1. **La donnée ne vaut rien** : seule compte sa réutilisation.
2. **La qualité des données est un résultat** plus qu'un prérequis
3. La donnée : **penser à la produire** autant qu'à la valoriser
4. La valeur n'attend pas le nombre des octets
5. Si je ne valorise pas mes données, **un autre le fera à ma place**
6. Le « big data » est un outil de **stratégie de la microdécision**
7. Il n'y a pas de projets data, mais une **vision data des projets**
8. La donnée n'est que **l'ombre de la réalité**
9. Valoriser la donnée est **un projet métier**, pas informatique

Un calendrier possible de formation continue

Exemple de calendrier



Programme proposé



Sem. 1
Sem. 2
Sem. 3
Sem. 4
Sem. 5
Sem. 6
Sem. 7
Sem. 8
Sem. 9
Sem. 10
Sem. 11
Sem. 12
Sem. 13
Sem. 14
Sem. 15
Sem. 16
Sem. 17
Sem. 18
Sem. 19
Sem. 20
Sem. 21
Sem. 22
Sem. 23
Sem. 24
Sem. 25

Entretiens préparatoires



Groupe 1



Groupe 2



Groupe 3



Groupe 4



Séminaire Data

1 Introduction : IA / Big Data / ML

→ Définition, enjeux, acteurs, limites

2 Enjeux pour l'assurance

→ Cartographie des besoins métiers potentiels

3 Environnements technique

→ Calculs parallélisés Vs distribués, mémoire et disque (Demos python, Spark)

4 Apprentissage supervisé

→ Random Forest / SVM, démo sous python (régression, classification)

5 Autres apprentissages

→ Non supervisé, renforcement, graphe, text (Knn, TF/IDF, A/B Testing)

6 Gestion d'un projet data

→ Agilité, boucle d'apprentissage, lecture de résultats, datavisualisation, suivi de projet

7 Cadrage stratégie data

→ Méthode de déclinaison des objectifs de l'entreprise pour aligner ses projets data

Un calendrier de formation continue : séminaire

- ▶ Un séminaire pour **marquer le temps de l'action**
- ▶ La performance de cette formation sera mesurée à l'aune des **actions que ces managers initieront à son issue**
- ▶ Les promotions de managers constituées ont aussi pour vocation de faire **générer des projets**
- ▶ Nous proposons de clôturer la formation par **une à deux journée(s) de séminaire présentant les projets à mettre en oeuvre**
- ▶ **Nous organisons et animons le séminaire, mais pas les éventuels coûts logistiques (réservation de salle, buffet, déplacements, etc.)**

Exemple de programme

Accueil
Rétrospective de la formation

Veille data

→ **Présentation des projets des promotions**

Déjeuner avec un
Chief Data Officer

→ Vote sur les 4 meilleurs projets
par promotion

**Engagement de la DG de mettre
en œuvre un projet sur les 4
présentés**

Remise des certificats

Soirée

Une formation reconnue aux data science

- ▶ Mémorandum est un **cabinet de conseil en data stratégie** qui focus sur le **métier avant tout usage de la data**
- ▶ Mémorandum est **organisme de Formation**, déclaré sous le numéro 32 60 03135 60 auprès de la préfecture de l'Oise
- ▶ Nous avons développé **des supports de formation (+ 1000 slides)** uniques et spécialisés pour R, Python, Spark qui mélangent **métier, machine learning et big data**
- ▶ Nos clients nous recommandent avec **une note de satisfaction de 95% chez IBM** sur les sessions 2016 / 2017 par exemple
- ▶ **Exemples d'entreprises ayant suivi la formation « Python for data science : de numpy a Mlib » (3 jours) en 2016/2017 :**



- ▶ **Exemples d'entreprises ayant été accompagnée sur une formation « Stratégie Big Data » :**





✓ Une écoute active des besoins

Après avoir défini les objectifs et les besoins en formation, nous élaborons **un programme sur-mesure** et adapté au profil des stagiaires définis.



✓ Une équipe de formateurs aux compétences spécifiques

Nos formateurs et consultants possèdent une compétence technique des métiers de la finance, de la banque, de l'assurance ou de la protection sociale.



✓ Une approche personnalisée

Pour adapter le programme de formation aux profils et niveaux des participants, nous pourrons : **préciser avec vous les parcours types par population**, via un test en ligne pour définir leurs besoins, ou laisser la possibilité aux collaborateurs de choisir ce qui les intéresse le plus.



✓ Une méthode pédagogique éprouvée

Basée sur l'**alternance d'apports théoriques et de mises en pratique immédiates**, notre méthode pédagogique favorise largement le jeu des questions / réponses ainsi que l'échange des points de vue et expériences.



✓ **Une base de capitalisation riche et actualisée**

Le cabinet s'appuie sur des outils et une base de capitalisation riche et actualisée qui lui permettent de disposer de **données de marché à jour** (benchmark...)



✓ **Des supports pédagogiques complets**

Une **documentation complète** est remise aux participants en début de stage. Elle constituera un précieux outil de travail et un document de référence aussi bien pendant qu'après la formation. Elle comprend tous les supports et exercices présentés pendant le stage.



✓ **Une évaluation de la formation et un suivi dans le temps**

À l'issue de chaque formation, un questionnaire est envoyé aux participants afin qu'ils puissent **évaluer la formation et l'intervenant**. L'objectif est de mesurer la valeur ajoutée de la formation et le degré de satisfaction des participants. Elle permet en outre d'apporter des modifications pour les sessions suivantes.

Au-delà des outils, penser les enjeux des big data

- ✓ Nous insistons toujours chez nos clients sur l'**importance de la compréhension des enjeux métiers**
- ✓ Il existe des **méthodes, techniques et outils pour la saisir**
- ✓ C'est ce que nous présentons dans trois modules centrés stratégies :

A. Découvrir où la data peut aider

- ✓ Le Document de Stratégie Data
- ✓ Le Framework d'analyse de la valeur attendue
- ✓ La matrice d'analyse impact / complexité
- ✓ L'analyse de la chaîne de valeur

B. La Création d'équipe data pour relayer au niveau opérationnel :

- ✓ Recrutement
- ✓ Montée en compétence
- ✓ Gestion du Turn-over

C. Devenir une organisation data-driven pour relayer au niveau Stratégique :

- ✓ Sponsors des initiatives
- ✓ Implications à tout niveau hiérarchique

Ces modules sont détaillés dans les slides suivantes

Module A : Comment découvrir où utiliser la data ?

Objectifs de la formation



- « Il n'y a pas de projet data, que des visions data des projets »

Public concerné

- Manipulateur de la data ✓
- Clients de projets BD&IA ✓
- Data-scientists ✓
- Niveau : Débutant

Format de session proposé

- Cours magistral de 4h00
- 20 personnes max
- 1 session avec exercices
- Nombre de formateurs : 1

Compétences ciblées

- Développer une vision data de ses projets quotidiens
- Développer un regard aiguisé sur la potentialité de la data dans l'amélioration des processus

Plan de formation

- **Jour 1 (4h00):**
 - Le Document de Stratégie Data
 - Le Framework d'analyse de la valeur attendue
 - La Matrice de confusion
 - La Matrice coût opportunité
 - Calcul de lift
 - La matrice d'analyse impact / complexité
 - L'analyse de la chaîne de valeur

Module B : La création d'équipe data

Objectifs de la formation



- Il n'y a pas de data-scientists, il n'y a que des équipes de data-science

Public concerné

- Manipulateur de la data ✓
- Clients de projets BD&IA ✓
- Data-scientists ✓
- Niveau : Débutant

Format de session proposé

- Cours magistral de 2h30
- 20 personnes max
- 1 session avec des exercices
- Nombre de formateurs : 1

Compétences ciblées

- Apprécier la nécessité du travail en collaboration Métier <-> Data-Scientist pour la réussite des projets data
- Comprendre comment créer des équipes gagnantes sur les enjeux data

Plan de formation

- **Jour 1 (2h30):**
 - *The mythical datascientists*
 - Quelles compétences pour une équipe data ?
 - Comment recruter un data-scientist ?
 - Quelle montée en compétence ?
 - Comment les retenir ?

Module C : Devenir une organisation data-driven

Objectifs de la formation



- «Pour devenir data-driven, il faut le décider »

Public concerné

- Manipulateur de la data ✓
- Clients de projets BD&IA ✓
- Data-scientists ✓
- Niveau : Débutant

Format de session proposé

- Cours magistral de 2h30
- 20 personnes max
- 1 session : cours magistral
- Nombre de formateurs : 1

Compétences ciblées

- Faire une analyse critique de ses propres modes de pensées, développer une introspection sur ses modes de décisions
- Etendre cette analyse au niveau de l'entreprise : comment les décisions sont-elles prises ?

Plan de formation

- **Jour 1 (2h30):**
 - Comment prend-on des décisions ?
 - Qu'est-ce qu'une organisation data-driven ?
 - Les rôles types : CDO / Data Steward / Digital Archivist ...
 - Exemple d'organisation de datalab

Programme personnalisable sur 4 cibles

	Qui ?	Quoi ?
1	Collaborateurs éloignés des questions data	Du Big Data pour quoi faire ? Le Big Data dans votre secteur. IA, Machine Learning et Datavisualisation
2	Managers business	Découvrir où la data peut aider Démystification du Machine Learning Introduction à la Data Visualisation Environnement technique Big Data Gestion de projets data GDPR
3	Utilisateurs de la data (statisticiens, webanalysts, responsables de reporting...)	Python pour le Machine Learning Data Science appliquée : text mining , Deep learning, Graphminning Data Visualisation - d3js Plateforme Big data
4	Datascientists junior	Faire émerger un besoin avec les métiers Piloter un projet Data avec les métiers Présenter un résultat aux métiers

Module 1 : Compréhension des enjeux data

Objectifs de la formation



- Appréhender les leviers de transformation de l'entreprise par la donnée
- Démystifier le machine learning

Public concerné

- Clients de projets BD&IA ✓
- Manipulateur de la data ✓
- Data-scientists ✓
- Niveau : Débutant

Format de session proposé

- 20 personnes max
- 2 sessions :
 - jour 1 : Cours magistral : 1h30
 - jour 2 : Ateliers uses cases internes : 1h30
- Nombre de formateurs : 1

Compétences ciblées

- Passer de la réflexion à la pratique : comment puis-je commencer chez moi dès demain ?
- Comprendre la méthodologie de travail d'un data-scientist

Plan de formation

- **Jour 1 (1h30):**
 - Que faire avec ses données :
 - Étapes de valorisation
 - Cadre d'analyse et d'exploration des usages
 - Découvertes de nouveaux univers de données
 - Le Machine Learning vu de l'intérieur :
 - Les algorithmes expliqués sans formules mathématiques
 - Comment piloter un projet data ?
 - Comment décoder le travail d'un data-scientist ?
 - Principes de gouvernance de projet
 - Fondamentaux du règlement européen de protection des données personnelles
 - Un devoir est donné en fin de session : à chacun de se projeter et trouver des uses cases
- **Jour 2 (1h30)**
 - Dépouillement et travail sur les uses cases proposés

Module 2 : Démystification du machine learning

Objectifs de la formation



- Comprendre et pratiquer les principaux algorithmes de machine learning

Public concerné

- Clients de projets BD&IA ✓
- Manipulateur de la data ✓
- Data-scientists ✓
- Niveau : Débutant

Format de session proposé

- Une matinée
- Session jusqu'à 20 personnes
- Nombre de Formateurs : 2
- Cours magistral : 1h
- Cas pratique : 2h

Compétences ciblées

- Comprendre le vocabulaire indispensable du Machine learning
- Savoir que faire lorsqu'on se retrouve face à de la data : par quelles étapes commencer ?
- Avoir une vision sur les outils du big data

Plan de formation

- **Formateur 1:**
 - Principes fondamentaux des algorithmes de ML
 - Apprendre à diagnostiquer et évaluer le travail de data-scientist
 - Exploration algorithmique, une exploration de l'intérieur, sans formule mathématique :
 - Régression linéaire pénalisée,
 - Des arbres aux forêts, boosting,
 - Réseaux neuronaux et deep learning
 - Décoder le travail d'un data-scientist et gérer un projet data
 - Démarche et ressources pour monter en compétences régulièrement
- **Formateur 2:**
 - Ecosystème big data :
 - s'y retrouver dans la jungle des outils et infrastructure

Module 3 : Data-Science appliquée

Objectifs de la formation



- Trois jours pour comprendre et pratiquer les principaux algorithmes de machine learning

Public concerné

- Clients de projets BD&IA ✘
- Manipulateurs de la data ✔
- Data-scientists ✔

- Niveau : Débutant

Format de session proposé

- Deux jours
 - Cours magistral : 7h
 - Travaux pratiques : 7h
- Session jusqu'à 12 personnes
- Nombre de Formateurs : 1

Compétences ciblées

- Connaître les sous jacents mathématiques du Machine Learning
- Avoir une compétence technique par la pratique (sous R ou Python)

Plan de formation

- **Jour 1:**
 - Rappel des principes fondamentaux des algorithmes de machine learning
 - Exploration algorithmique (théorie et mise en pratique) :
 - Régression linéaire,
 - Sélection de feature, Régularisation ridge / Lasso, Elastic net

- **Jour 2:**
 - Des arbres aux forêts (2 sessions CART, Random Forest, GBM, Xgboost)
 - Apprentissage non supervisé : PCA, TSNE, factorisation, hypothèse de bas rang
 - SVM et méthodes à noyaux

Module 4 : Introduction à la Data Visualisation

Objectifs de la formation



- Comprendre les enjeux et les possibilités offertes par la datavisualisation

Public concerné

- Clients de projets BD&IA ✓
- Manipulateur de la data ✓
- Data-scientists ✓

- Niveau : Débutant

Format de session proposé

- Une matinée
- Cours magistral : 1h30
- Session jusqu'à 20 personnes
- Nombre de Formateurs : 1

Compétences ciblées

- Savoir éclaircir ses présentations en supprimant le superflu
- Apprendre à communiquer au métier : passer des résultats de maths aux enjeux business
- Connaître le nom des technologies existantes

Plan de formation

- **Concepts de la datavisualisation**
 - Règles de Tufte

- **Présentation d'exemple**
 - Historique (Marche de russie)
 - Modernes (vidéo de Hans Rosling).

- **Listing des technologies :**
 - Web (HTML, CSS, AJAX, SVG, d3.js)
 - Programmative (Python matplotlib, bokeh, plotly.... R ggplot)

Module 5 : Environnement technique Big Data

Objectifs de la formation



- Se repérer dans les technologies Big Data majeures

Public concerné

- Clients de projets BD&IA ✓
- Manipulateur de la data ✓
- Data-scientists ✓

- Niveau : Moyen / Avancé

Format de session proposé

- Cours magistral : 3h00
- Une matinée
- Session jusqu'à 20 personnes
- Nombre de Formateurs : 1

Compétences ciblées

- Comprendre le fonctionnement des calculs distribués
- Connaître les noms des principales technologies, et leurs différences

Plan de formation

- **Notions de réseau :**
 - Routeurs, ports, serveurs, ssh

- **Présentation de l'écosystème Hadoop v1**
 - Hive, Pig, Yarn, Giraf, Mahout...
 - et v2 : Hortonworks

- **Émergence de Spark comme remplaçant d'Hadoop**

- **Présentation des composants d'une infra moderne :**
 - Mesos / Marathon / Zookeeper
 - Kappa Architecture

- **Exemple d'une architecture d'un site web, et récolte de logs sur Kafka.**

Module 6.1 : Python / PySpark – partie 1

Objectifs de la formation



- Apprendre Python

Public concerné

- Clients de projets BD&IA ✓
- Manipulateur de la data ✓
- Data-scientists ✓
- Niveau : Débutant

Format de session proposé

- Cours magistral et TP: 7h00
- 1 journée
- Session jusqu'à 12 personnes
- Nombre de Formateurs : 1

Compétences ciblées

- Savoir coder en python

Plan de formation

- **Une journée pour acquérir les bases de python**
 - Environnements de développement
 - Syntaxes
 - Variables
 - Classes
 - Exercices

Module 6.2 : Python / PySpark – partie 2

Objectifs de la formation



- Apprendre Python pour le Machine Learning : Matplotlib et Scikit-Learn

Public concerné

- Clients de projets BD&IA ✓
- Manipulateur de la data ✓
- Data-scientists ✓

- Niveau : Intermédiaire

Format de session proposé

- Cours magistral et TP: 7h00 / jour
- 2 journées
- Session jusqu'à 12 personnes
- Nombre de Formateurs : 1

Compétences ciblées

- Savoir utiliser les bibliothèques de machine learning de Python (numpy, pandas, scikit-learn...)
- Savoir faire un algorithme prédictif de régression et de classification

Plan de formation

Deux journées pour découvrir Python pour le ML les bases de python

- **Jour 1:**
 - Introduction à numpy, pandas
 - Rappel de Machine Learning
 - Découverte de Matplotlib
 - Découverte de Scikit Learn
 - Exercices

- **Jour 2**
 - Rappel de python
 - Rappel de Machine Learning
 - Exercices Scikit-Learn / Matplotlib

Module 6.3 : Python / PySpark – partie 3

Objectifs de la formation



- Apprendre Spark pour le Machine Learning

Public concerné

- Clients de projets BD&IA x
- Manipulateurs de la data x
- Data-scientists ✓

- Niveau : Intermédiaire / avancé

Format de session proposé

- Cours magistral et TP: 7h00 / jour
- 3 journées
- Session jusqu'à 12 personnes
- Nombre de Formateurs : 1

Compétences ciblées

- Savoir monter un cluster spark et hadoop « from scratch »
- Comprendre leur fonctionnement et les points capitaux pour leur performance
- Savoir faire un modèle de ML en spark

Plan de formation

Trois journées pour découvrir Python pour le ML les bases de python

- **Jour 1:**
 - Introduction aux big data (1h30)
 - Introduction à Spark
 - Mise en place d'un cluster Spark
- **Jour 2**
 - Introduction à Hadoop
 - Mise en place d'un cluster HDFS
 - Communication entre Spark et HDFS
- **Jour 3**
 - Introduction au Machine Learning
 - Introduction à MLlib
 - Modélisation de données stockées sur HDFS à travers MLlib

Module 7 : Datavisualisation

Objectifs de la formation



- Apprendre d3js

Public concerné

- Clients de projets BD&IA **x**
- Manipulateur de la data **x**
- Data-scientists **✓**

- Niveau : Avancé

Format de session proposé

- Cours magistral : 1h30
- TP: 3h30
- 1 après-midi
- Session jusqu'à 12 personnes
- Nombre de Formateurs : 1

Compétences ciblées

- Savoir faire une datavisualisation interactive sur une page web avec d3js

Plan de formation

- **Une journée pour découvrir la datavisualisation et d3js**
 - Rappels théoriques sur la datavisualisation
 - Présentation des technologies web (AJAX, d3js)
 - Exercice : tracer le cac40 en d3.js

Module 8.1 : Théorie du text mining - débutant

Objectifs de la formation



- Apprendre des concepts mathématiques de manipulation de texte

Public concerné

- Clients de projets BD&IA ✘
- Manipulateurs de la data ✔
- Data-scientists ✔

- Niveau : Avancé

Format de session proposé

- Cours magistral : 1h30
- 1 après-midi
- Session jusqu'à 20 personnes
- Nombre de Formateurs : 1

Compétences ciblées

- Connaître les principes théoriques du text mining

Plan de formation

- **Présentation de lois mathématiques de text-mining :**
 - Loi de heaps
 - Loi de zipf
 - Lemmatisation
 - Racinisation
 - TF IDF
 - Rntity recognition
 - ...

Module 8.2 : Text Mining théorie et pratique

Objectifs de la formation



- Apprendre des concepts mathématiques de manipulation de texte

Public concerné

- Clients de projets BD&IA ✘
- Manipulateurs de la data ✔
- Data-scientists ✔

- Niveau : Avancé

Format de session proposé

- Cours magistral : 1h30
- TP: 3h30
- 1 après-midi
- Session jusqu'à 12 personnes
- Nombre de Formateurs : 1

Compétences ciblées

- Connaître les principes théoriques du Text Mining
- Avoir une pratique technique du Text Mining

Plan de formation

- **Présentation de lois mathématiques de Text Mining :**
 - Loi de heaps
 - Loi de zipf
 - Lemmatisation
 - Racinisation
 - TF IDF
 - Entity recognition
 - ...

- **Exercices en R ou Python**

Module 9 : Deep Learning

Objectifs de la formation



- S'immerger dans les concepts théoriques du Deep Learning

Public concerné

- Clients de projets BD&IA ✘
- Manipulateurs de la data ✔
- Data-scientists ✔

- Niveau : Débutant

Format de session proposé

- Cours magistral : 1h30
- 1 matinée
- Session jusqu'à 20 personnes
- Nombre de Formateurs : 1

Compétences ciblées

- Connaître les principes théoriques du Deep learning
- Connaître leurs cas d'usage et de non-usage

Plan de formation

- **Historique du Deep Learning**
 - Perceptron

- **Nouvelle ère :**
 - Deep feed forward network
 - Fonction d'activation
 - Back Propagation
 - Régularisation
 - Early stopping

- **Architecture des réseaux de neurones**

Module 10 : Gestion de projet data

Objectifs de la formation



- Trois jours pour comprendre et pratiquer les principaux algorithmes de machine learning

Public concerné

- Clients de projets BD&IA ✓
- Manipulateur de la data ✓
- Data-scientists ✓
- Niveau : Débutant

Format de session proposé

- Cours magistral : 1h
- Cas pratique : 2h
- Une matinée
- Session jusqu'à 20 personnes
- Nombre de Formateurs : 1

Compétences ciblées

- Comprendre les méthodes de travail des data scientists
- Comprendre l'importance de la *feature engineering* et de la connaissance métier
- Avoir des outils pour traduire les résultats mathématiques en output métier

Plan de formation

- **Jour 1:** Gestion d'un projet data : quelle est la vraie valeur ajoutée du data-scientist
 - Les écueils fréquents
 - « Back to the feature »
 - Restituer un résultat compréhensible par le métier
 - Gestion Scrum

Module 11 : Fondamentaux de l'IA

Objectifs de la formation



- 3 heures pour comprendre ce qui se cache vraiment derrière le « buzzword : intelligence artificielle »

Public concerné

- Clients de projets BD&IA ✓
- Manipulateur de la data ✓
- Data-scientists ✓

- Niveau : Débutant

Format de session proposé

- Une matinée
- Cours magistral : 3h
- Nombre de Formateurs : 1 à 2
- Sessions jusqu'à 20 personnes

Compétences ciblées

- Acquérir une vision critique sur un buzz word, et se préparer pour sa réalité à venir ...

Plan de formation

- **QUOI ?**
 - Qu'est-ce que l'intelligence artificielle ?
 - Les différentes disciplines de l'IA - qu'est-ce qui en est et qu'est qui n'en est pas ?
- **QUAND ?**
 - Des années 50 à aujourd'hui, pourquoi en parlons-nous tant maintenant ?
- **COMBIEN ?**
 - Grands chiffres et perspectives ?
 - Comment l'IA est en passe de devenir la 4^e révolution industrielle ?
- **QUI ?**
 - Les acteurs de l'IA sur le marché ?
 - Les acteurs de l'IA dans l'entreprise ?
- **COMMENT ?**
 - Comment le marché approche l'IA ?
 - Comment l'IA impacte la chaîne de valeur ?
 - Comment mener un projet d'IA ?
- **Cas pratique** : Redessinons la chaîne de valeur d'une activité via l'IA

Module 12 : Graphmining

Objectifs de la formation



- Avec les réseaux sociaux la science du graph a complètement été révolutionnée, s'appropriier les concepts et la puissance potentielle

Public concerné

- Clients de projets BD&IA ✓
- Manipulateur de la data ✓
- Data-scientists ✓
- Niveau : Débutant

Format de session proposé

- Cours magistral : 3h & Cas pratique : 1h
- Une matinée
- Nombre de Formateurs : 1 à 2
- Session jusqu'à 20 personnes

Compétences ciblées

- Connaissance des algorithmes d'analyse des graphs
- Approche de Neo4j
- Compréhension de leurs cas d'usages

Plan de formation

- **Réseau et graphes**
 - Qu'est-ce qu'un graphe ? Notions et terminologies
 - Graphs et Web / Réseaux sociaux
- **Les algorithmes de graphes**
 - Matrice d'adjacence
 - Graphes spéciaux (arbres, DAG)
 - La traversée des graphs (profondeur, largeur...)
- **Impact sur les bases de données**
 - Du MCD classique aux bases graphs
 - Présentation de Neo4J

Module 13 : Plateforme Big Data

Objectifs de la formation



- Découvrir une distribution big data, penser le temps réel et la kappa architecture

Public concerné

- Clients de projets BD&IA ✘
- Manipulateurs de la data ✘
- Data-scientists ✔
- Niveau : avancé

Format de session proposé

- Cours magistral : 4h
- Une matinée
- Nombre de Formateurs : 1 à 2
- Session jusqu'à 20 personnes

Compétences ciblées

- Connaître les principes théoriques sous-jacents du Big Data
- Comprendre les enjeux en termes d'infrastructure

Plan de formation

- **Présentation d'Hortonworks**
 - Quels outils
 - Quel usage
- **Approfondissement sur Hadoop**
 - Fonctionnement interne
 - Symbiose avec Spark
- **Approfondissement sur Spark**
 - Fonctionnement interne
 - Distribution de calcul
 - Temps réel
- **Présentation de Kafka**
 - Cas d'usage
 - Logique de routage de messages
 - Intégration avec Spark

Module 14 : GDPR

Objectifs de la formation



- Comprendre les enjeux du GDPR

Public concerné

- Clients de projets BD&IA ✓
- Manipulateur de la data ✓
- Data-scientists ✓
- Niveau : Débutant

Format de session proposé

- Cours magistral : 4h
- Une matinée
- Nombre de Formateurs : 1 à 2
- Session jusqu'à 20 personnes

Compétences ciblées

- Connaissance des enjeux de la réglementation européenne

Plan de formation

- **Histoire de la CNIL**
 - Les réglementations
 - L'évolution des normes
 - La modification du marché
- **Le droit des données à l'étranger**
 - Amérique Vs Europe
 - Harmonisation des politiques européennes
- **Le GDPR**
 - Enjeux
 - Application
 - Sanctions

Quelques points logistiques

- ✓ **Fiabilité :**
 - ✓ Nos **formateurs** sont :
 - ✓ **diplômés des plus grandes écoles** : MS Big Data de Télécom Paris, École Centrale, ESCP, CNAM....
 - ✓ **Des professionnels de la data** : Cabinet Mémorandum, Directeur du MS « Assurance, Actuariat et Big Data » de l'ESILV
- ✓ **Facilité et confort de fonctionnement**
 - ✓ **Nous venons dans vos locaux**, nous vous proposons des locaux à la Défense au besoin (locaux de l'ESILV)
- ✓ **Capacité d'évolution**
 - ✓ Les **contenus de cours sont adaptables** lors d'une prochaine discussion.
- ✓ **Références**
 - ✓ Formation continue : **Axa, IBM, Orange, Crédit Agricole, Alstom, Matmut...**
 - ✓ Formation initiale : ESILV, Ecole Supérieur de l'Assurance, EMGrenoble, ESCP, Université Paris XIII, ISG Business School...
- ✓ **Moyens mis en œuvre et qualité de l'organisation proposée**
 - ✓ Moyens : **Corps Enseignant expérimenté** et diplômé BAC +5 minimum des grandes écoles
 - ✓ Qualité : Très bonnes notations de nos formations – **95% de satisfaction chez IBM**



Denis Oblin

Associé – Expert Data Science & business analysis

En bref

Parcours

1. Vertone (Manager)
2. Groupama (Directeur marché des particuliers)
3. Mémoire (Associé)

Formation et expérience

- Centrale Nantes 1994
- Telecom ParisTech (mastère Big data 2014)

Secteurs

1. Banques et Assurances
2. Télécom
3. E commerce

Expertises

- Marketing client
- Gestion de réseau commercial
- Machine learning – R

Contacts

- Denis.oblin@memorandum.pro
- 06.71.62.74.92

Exemples de projets réalisés

Projet	Client	Détail
Déploiement du datalab	Matmut	<ul style="list-style-type: none"> • Emergence des projets data dans toutes les directions de l'entreprise • Mise en place une gouvernance des ressources datascience • Animation des data scientists internes : algorithmique / gestion de projet / restitution
Coaching de l'équipe Datascience	AXA Life Invest	<ul style="list-style-type: none"> • Animation du pool Datalab sur 6 mois • Aide au cadrage et à la réalisations de projets opérationnels (pilotage réseaux commerciaux, évolution produit, analyse client) • Formation ad hoc au fil de l'eau des besoins rencontrés sur les projets
POC Datascience avec transfert de compétences	Randstad	<ul style="list-style-type: none"> • Diagnostic de performance du réseau commercial, croisant une centaine de tables • Formalisation du diagnostic opérationnel • Transfert de compétences aux équipes statistiques internes au fil de l'eau du projet
Enseignement du module data science 5é année ingénieur et mastère assurance actuariat et Big Data	ESILV	<ul style="list-style-type: none"> • 40 heures de cours et TD dédiés au machine learning, théorie et mise en pratique. 10 algorithmes pratiqués • 3è année d'enseignement en cours de déploiement
Valorisation de 1M de dossiers médicaux	Terre des Hommes Suisse	<ul style="list-style-type: none"> • Déploiement de 5 cas d'usages opérationnels basés sur l'exploitation des dossiers médicaux au Burkina Faso (gestion de site, ..) • Formation des équipes locales au fil de l'eau
R&D	Welcome Track	<ul style="list-style-type: none"> • Développement depuis 3 ans des travaux de recherche et développement : évolution des services basés sur l'exploitation des données collectées (transport colis e commerce)



Romain Jouin

Associé – Expert Data Science & business analysis

En bref

Parcours

1. Alcatel (Section mobile)
2. Toshiba (Cloud Computing)
3. Quinten (Data-Scientist)
4. Mémoire (Associé)

Formation et expérience

- Telecom Sud Paris 2006
- ESCP (MS) 2010
- Telecom ParisTech (mastère Big data 2014)

Secteurs

1. Télécom
2. Cloud
3. E commerce

Expertises

- Développement informatique
- Machine learning – python
- Big Data (Spark / HDFS)

Contacts

- romain.jouin@memorandum.pro
- 06.52.86.87.30

Exemples de projets réalisés

Projet	Client	Détail
Déploiement du datalab	Matmut	<ul style="list-style-type: none"> • Supervision de l'infrastructure fonctionnelle • Mise en place une gouvernance des ressources datascience • Animation des data scientists internes : algorithmique / gestion de projet / restitution
Formateur Spark / Python (+150h)	m2i / LearnQuest / ambient-it / Opcadia	<ul style="list-style-type: none"> • Formations introductives génériques aux big data (EDF, CA) • Formateur IBM : « De Numpy à MLib – Python pour la datascience » : 7 jours en 2016 et 2017 • Formateur CGI : « PySpark pour la datascience » – 3 jours
POC Datascience avec transfert de compétences	Randstad	<ul style="list-style-type: none"> • Diagnostic de performance du réseau commercial, croisant une centaine de tables • Formalisation du diagnostic opérationnel • Transfert de compétences aux équipes statistiques internes au fil de l'eau du projet
Directeur du mastère « Assurance, Actuariat et Big Data »	ESILV	<ul style="list-style-type: none"> • 70 heures de cours et TD dédiés aux Big Data, théorie et mise en pratique. • Gestion de 22 intervenants extérieurs • 3^e année d'enseignement en cours de déploiement
R&D : analyse de satisfaction client	Satisfactory	<ul style="list-style-type: none"> • Aider à trouver des points significatifs dans des séries temporelles de satisfaction client en fonction des contextes d'exécution des services
R&D : outil de prévision de Tarif	Loc'Nacelle	<ul style="list-style-type: none"> • Développement d'un algorithme d'automatisation de pricing selon le client, les dates et les disponibilités machines. • Mise en production d'un API prédictive de la flexibilité au Tarif.