

Machine Learnia Ltd

Cambridge
United Kingdom

Formation ML-PRO de Machine Learnia

Formateur: Guillaume Saint-Cirgue

CURSUS	3
1. Les Fondations	3
2. Spécialisation Mathématique	3
3. Spécialisation Machine Learning	3
4. Spécialisation Deep Learning	4
5. Spécialisation Gestion de Projet	5
6. Spécialisation Data Engineering	6
7. Spécialisation Cloud - Déployer et maintenir vos modèles de ML dans le monde réel	6
8. Spécialisation Apprentissage par Renforcement	6
9. Spécialisation IA Générative	6

CURSUS

1. Les Fondations

- Le raisonnement scientifique, l'approche de travail et la méthodologie pour résoudre des problèmes de Data Science.
- Prise en Main machine: Linux, gestion des Venvs, Terminal, Bash, zsh, formation VSCode, Pycharm, VIM
- Formation GIT et github, mise en pratique
- Formation SQL, SQLite, mise en pratique
- Formation sklearn, mise en pratique
- Formation pandas, mise en pratique
- Formation matplotlib, mise en pratique
- Formation numpy, mise en pratique
- Formation streamlit, mise en pratique

2. Spécialisation Mathématique

- Formation Statistiques et probabilités: Probabilités, Statistiques, Lois de probabilités, test d'hypothèse, échantillonnage, estimation, plan d'expérience, exercices et mise en pratique avec numpy et scipy.
- Formation Algèbre linéaire, matrices, valeurs propres et vecteurs propres, décomposition en valeur singulière, exercices, mise en pratique, et création d'applications de visualisation avec MANIM
- Formation Optimisation: Algorithmes de minimisation, programmation linéaire, théorie des graphes, exercices et mise en pratique avec numpy et scipy
- Formation Analyse fréquentielle et Transformée de Fourier: nombres complexe, trigonométrie, FFT, mise en pratique avec numpy et scipy

3. Spécialisation Machine Learning

- Cours détaillé sur le fonctionnement des algorithmes de Machine Learning:
 - Les Modèles Linéaires: Régression linéaire, Régression Logistique, Méthodes LARS, Méthode OMP, etc
 - Les Arbres de Décision.
 - Les K-Nearest Neighbors

- Les modèles ensemblistes: Random Forest, AdaBoost, Gradient Tree Boosting, etc
- Les modèles gaussiens
- Les Support Vector Machines
- Les modèles de Clustering: K-Means, DBSCAN, Hierarchical, Optics, etc.
- Les méthodes de réduction de dimension: T-SNE, PCA, UMAP, etc.
- Les Méthodes de détection d'anomalie
- Formation sur les algorithmes de série temporelle, mise en pratique avec Statsmodels

4. Spécialisation Deep Learning

- Un rappel sur bases du Deep Learning (déjà vu sur YouTube via ma série de vidéos) un résumé pratique permettant de consolider et d'affirmer vos compétences dans ce domaine.
- La suite et fin de cette série, exposant les méthodes de deep learning plus avancées: Fonctions d'activations, Méthodes d'initialisation (Xavier et Bengio), Batch Normalisation, Multi-classification, Dropout, Algorithmes d'optimisation plus poussées (Adam, RMSProp, etc), Batch Learning, etc.
- Un passage à la pratique, avec un cours complet sur Keras et TensorFlow, pour être capable de développer des réseaux de neurones dans de vrais projets. Utilisation de CUDA et des fonctionnalités avancées des frameworks
- Un cours équivalent sur l'utilisation de Pytorch, le framework d'excellence utilisé dans la recherche et de nombreuses startups.
- Un cours approfondi sur les autres grands domaines du Deep Learning:
 - Les CNNs pour la vision par ordinateur: comprendre et maîtriser les bases, puis les différentes architectures répandues (VGG, AlexNet, ResNet, MobileNet etc.) Ainsi que les modèles de détection, comme YOLO, et de segmentation, comme SAM. Comment faire du Transfer learning. Des exercices et des cas pratiques sur Tensorflow et Pytorch.
 - Pour pratiquer: des projets de vision par ordinateur: Détection d'objets, segmentation d'image, etc.

- Les RNNs pour le traitement de séquence et le traitement du langage: maîtriser les bases, puis les différentes architectures (LSTM, GRU).
 - Pour pratiquer: des projets concrets de séries temporelles, et sur le traitement du langage naturel (NLP) comprenant l'analyse de sentiment.
- Les Auto-encodeurs, les GANs, les Transformers et les autres techniques modernes à connaître en 2024+. Comment fonctionne chaque méthode et comment les implémenter avec Tensorflow et Pytorch

Pour finir ce module, une série de projets pratiques

5. Spécialisation Gestion de Projet

- Gestion de projets de Machine Learning, checklist et roadmap d'un projet
- Formation visualisation de données et EDA:
 - Méthode OPEC
 - Regles de la dataviz
 - Storytelling with data
 - Nombreuses études de cas
- Formation Preprocessing:
 - Comment sélectionner les variables les plus pertinentes
 - Comment nettoyer vos données pour obtenir un ensemble propre et optimal pour votre modèle de Machine Learning.
 - Comment gérer les variables catégorielles et numériques
 - Quelles techniques de normalisation utiliser
 - Feature engineering
 - Comment aborder les problèmes de classification déséquilibrée, et quels sont les outils pour y remédier.
 - La maîtrise et la compréhension des 29 opérations de pre-processing offertes par Sklearn.
- Formation Evaluation:
 - Processus d'évaluation diagnostique, error analysis
 - Métriques de classification et de régression, exercices et mise en pratique
 - Quantification des incertitudes
- Formation DEVOPS et MLOPS
 - Pratique avec MLFlow
- Regulations IA - Ethique, GDPR

6. Spécialisation Data Engineering

- Concepts du Data Engineering
- Formation Spark
- Formation Docker
- Formation K8s
- Formation Terraform
- Formation Postgre / DuckDB
- Formation Airflow

7. Spécialisation Cloud - Déployer et maintenir vos modèles de ML dans le monde réel

- Formation aux concepts du Cloud
- Formation AWS / Azure / GCP: Comment développer, Déployer, et Maintenir vos modèles de Machine Learning sur ces différentes plateformes
- Projet: Créer un SaaS de Machine Learning.

8. Spécialisation Apprentissage par Renforcement

- Formation Apprentissage par renforcement: Chaînes de markov, Q-Learning, Monte Carlo
- Formation Algorithmes Génétiques
- Projets

9. Spécialisation IA Générative

- Formations LLMs
 - Introduction aux architectures des LLMs
 - RAGs
 - Utilisation des modèles Open Source et déploiement
- Formation Génération d'image

- Introduction aux architectures de diffusion
- Utilisation des modèles Open Source et déploiement