

Machine Larnia Ltd

Cambridge
United Kingdom

Formation ML-PRO de Machine Larnia

Formateur: Guillaume Saint-Cirgue

CURSUS

1. Les Fondations	3
2. Spécialisation Mathématique	5
3. Spécialisation Machine Learning	6
4. Spécialisation Deep Learning	7
5. Spécialisation Gestion de Projet	8
6. Spécialisation Data Engineering	9
7. Spécialisation Cloud - Déployer et maintenir vos modèles de ML dans le monde réel	10
8. Spécialisation Apprentissage par Renforcement	11
9. Spécialisation IA Générative	12

1. Les Fondations

- **Prise en main de sa machine:**
 - Comprendre l'architecture d'un ordinateur
 - Guides d'installation des outils pour Linux, MacOS et Windows
 - Utilisation de WSL2 avec Windows
 - Utiliser VIM
 - Utiliser le terminal ZSH et les plugins de OH-MY-ZSH
 - Gestion des environnements Python et des projets:
 - Anaconda
 - Miniconda
 - Pyenv et Poetry
 - UV
 - Utiliser VSCode et les extensions importantes
 - Les bases de PyCharm
- Les **frameworks** de la data science:
 - Matplotlib et seaborn
 - Numpy
 - Pandas
 - Scipy
 - Scikit-learn
 - Keras
 - Streamlit
- Formation **GIT** et **github**
- Formation **SQL**, **SQLite3** et projet avec Streamlit
 - SQL et les différents SGBD
 - Les bases de données relationnelles
 - Les opérations de base en SQL
- Le **raisonnement scientifique**, l'**approche de travail** et la **méthodologie** pour résoudre des problèmes de Data Science.
 - Analyse univariée et multivariée
 - Formuler et tester des hypothèses (Introduction aux tests)
 - Analyse fondamentale
 - Pipeline de développement d'un modèle de machine learning
 - Introduction au preprocessing
 - Introduction à la modélisation
 - Les branches du machine learning, la carte de l'IA, etc.
 - Introduction à l'évaluation : diagnostiquer un modèle, overfitting et underfitting, etc.

- **Formation Python:**

- Les types de données, les opérations et expressions
- Les conditions et structures match-case
- Les boucles
- Les listes, tuples, ensembles, dictionnaires
- La gestion des fichiers
- Les fonctions
- Écrire des docstrings et structurer son code
- Les modules de base (math, datetime, os, glob, string, random, csv, JSON)
- La POO
- Les modules collections, itertools, re
- Les exceptions et le logging
- Les lambdas et outils fonctionnels
- La récursivité
- Créer des packages
- Les décorateurs
- Les générateurs
- Programmation asynchrone et multi-threading
- Les modules functools et sys
- Exercices et projets

2. Spécialisation Mathématique

- Formation **Statistiques et probabilités**:
 - Statistiques descriptives
 - Lois de probabilités
 - Tests d'hypothèse
 - Échantillonnage
 - Estimation
 - Plan d'expérience
 - Exercices et mise en pratique avec numpy et scipy
- Prise en main de **Manim**, la librairie python utilisée pour faire des animations mathématiques
- Formation **Algèbre linéaire**:
 - Les ensembles
 - Les vecteurs et matrices
 - Espaces vectoriels
 - Applications linéaires
 - Le déterminant
 - L'inverse d'une matrice
 - Les valeurs propres et vecteurs propres
 - Polynôme caractéristique
 - Diagonalisation et trigonalisation
 - Décomposition en valeurs singulières
 - Exercices, mise en pratique, et 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
 - Ridge, Lasso, ElasticNet
 - 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 support vector classifiers (Linéaires et à noyaux)
 - Les support vector regressors
 - Les nuSVC
 - **Naive Bayes**
 - Les modèles de **Clustering**:
 - K-Means
 - DBSCAN
 - Hierarchical
 - Optics
 - Mean shift
 - etc.
 - Les méthodes de **réduction de dimension**:
 - T-SNE
 - PCA
 - UMAP
 - etc.
- Les Méthodes de **détection d'anomalies**
- Formation sur les algorithmes de **série temporelle**, mise en pratique avec Statsmodels

4. Spécialisation Deep Learning

- Un rappel sur les **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
 - Les différentes architectures répandues (VGG, AlexNet, ResNet, MobileNet etc.)
 - Les modèles de détection comme YOLO et de segmentation comme SAM
 - Faire du Transfer learning
 - 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 et 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-encoders (AE)
 - Les AE convolutionnels
 - Les AE variationnels
 - Les AE de débruitage
 - Les GANs
 - Les Transformers
- 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 pré-processing offertes par Sklearn
- Formation **Évaluation**:
 - 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
- Formation sur la **Regulation IA**: Ethical-AI, RGPD, EU AI Act

6. Spécialisation Data Engineering

- Formation Architecture
- Formation Pipelines ETL
- Formation Spark
- Formation Docker
- Formation Postgre
- Formation DuckDB
- Formation Airflow

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

- Introduction aux concepts du **Cloud**
- Formation **AWS** : 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
- Les **Algorithmes Génétiques**
- Projets

9. Spécialisation IA Générative

- Formation **LLMs**
 - Introduction aux architectures des LLMs
 - RAGs
 - Utilisation des modèles Open Source et déploiement
- Formation **Génération d'images**
 - Introduction aux architectures de diffusion
 - Utilisation des modèles Open Source et déploiement
- Formation **AgenticAI**