

Ceinture Blanche - Novice

Objectifs : Introduction au machine learning et aux outils de base.

- **Python :** Connaissance de la syntaxe de base de Python, utilisation des boucles, conditions, fonctions, et manipulation de types de données courants (listes, dictionnaires). Maîtrise de la Programmation Orientée Objet
- **Mathématiques :** Comprendre les concepts de base des statistiques descriptives (moyenne, écart-type) et des probabilités élémentaires.
- **Machine Learning :** Connaître les principes de base du machine learning: Apprentissage supervisé, Apprentissage non-supervisé, et Apprentissage par renforcement, et leur fonctionnement de haut niveau.
- **Outils :** Installation de Python, maîtrise du terminal et configuration de VSCode

Accomplissements :

- ☐ Vous avez installé Python sur votre ordinateur et vous savez jongler d'une version à l'autre en manipulant les environnements virtuels
- ☐ En terme de programmation Python: vous êtes capable de créer des fonctions, des classes, et vous savez manipuler des listes et des dictionnaires afin de résoudre des petits problèmes python, par exemple un algorithme de travelling salesman
- ☐ Vous savez vous servir de votre terminal et votre shell (powershell ou bien Bash / oh-my-zsh) et vous l'utilisez au quotidien lorsqu'il s'agit de travailler (ouvrir des dossier, lancer des script, coder, etc)
- ☐ Vous êtes capable de décrire les grandes étapes d'un projet de Machine Learning: l'exploration de données, le preprocessing, la modélisation, et l'évaluation de modèles.

Ceinture Jaune - Développeur initié

Objectifs : Devenir un développeur initié et autonome, savoir manipuler des données et utiliser les bibliothèques de base en machine learning pour réaliser des projets simples:

- **Python :** Maîtrise des bibliothèques de base:
 - Numpy et Pandas pour la manipulation des données
 - Matplotlib et Seaborn pour la création de graphique
 - Sklearn pour le pre-processing, la création de pipelines, l'entraînement de modèles, la recherche d'hyper-parametres, et l'évaluation finale avec les bonnes métriques de bases (telles que MSE et Accuracy)
- **Machine Learning :** Vous connaissez et maîtrisez au moins les algorithmes suivants, et vous pouvez les implémenter avec Scikit-Learn:
 - LinearRegression
 - LogisticRegression
 - KNNClassifier
 - DecisionTreeClassifier
- **SQL :** Connaissances de base en SQL pour récupérer des données, utilisation des commandes SELECT, WHERE, JOIN.
- **Maîtrise de Git**

Accomplissements :

- ☐ Vous êtes désormais un développeur Python autonome: vous savez lire et comprendre vos erreurs (Exception) de vous-meme, et vous savez vous servir de votre IDE (par exemple VSCode) pour débbugger vos projets python.
- ☐ Vous savez mener une analyse de données et vous pouvez explorer un dataset avec Pandas et Matplotlib.
- ☐ Vous avez pu créer des dashboards simples avec Streamlit, intégrant des filtres et des interactions basiques
- ☐ Vous avez pu Implémenter des modèles de machine learning de A-Z pour la régression et la classification
 - Preprocessing
 - Selection du modele
 - Validation croisée et recherche d'hyper-paramètre
 - Évaluation
- ☐ Vous avez réaliséé un petit projet mêlant SQL et Pandas, par exemple, extraire des données d'une base SQL, les charger en Pandas, y appliquer des transformations, et sauvegarder les résultats finaux
- ☐ Vous avez un compte Github ou Gitlab, et vous utilisez git quotidiennement quand vous codez, et cela se voit sur votre profil github !

Ceinture Orange - Mathématiques Avancées

Objectifs : Développer une compréhension approfondie des mathématiques nécessaires au machine learning.

- **Algèbre Linéaire :** Compréhension des concepts de vecteurs, matrices, opérations matricielles (produit matriciel, déterminant, inverse). Compréhension des valeurs propres et vecteurs propres. Etre capable de décomposer une matrice avec les algorithmes de SVD, et d'utiliser cela pour résoudre des problèmes
- **Statistiques et probabilités :** Maîtrise des distributions de probabilité, calculs de variance et covariance, théorèmes fondamentaux de la probabilité. Maîtrise des tests d'hypothèse et calcul à la main.
- **Algorithmes d'optimisation:** Connaissance et maîtrise des familles d'algorithmes d'optimisation présentes dans scipy

Accomplissements :

- ☐ Vous avez pu résoudre des exercices pratiques à la main et sur ordinateur:
 - ☐ de test d'hypothèses: Student (unilatéral et bilatéral) Chi2, ANOVA, Fisher...
 - ☐ Estimer les paramètres d'une loi de probabilité
 - ☐ pouvoir utiliser la loi Normale, loi de Poisson, Loi Binomiale, Loi Hypergéométrique, pour résoudre des problèmes d'entreprise.
 - ☐ Vous êtes capable d'estimer l'incertitude autour de la performance d'un modèle de régression et d'un modèle de classification
- ☐ Projet d'algèbre linéaire:
 - ☐ Réaliser un PCA uniquement avec Numpy
 - ☐ Réaliser une Régression Linéaire avec l'algorithme du SVD.

Ceinture Verte - Maîtrise des Algorithmes de Machine Learning

Objectifs : Comprendre en profondeur les algorithmes classiques du machine learning.

- **Algorithmes :** Maîtrise de tous les algorithmes de machine learning (forêts aléatoires, SVM, arbres de décision, k-means, ...) avec une compréhension approfondie de leurs hyperparamètres, et comment les régler pour obtenir les meilleures performances.
- **Maîtrise de toutes les techniques de pre-processing** (feature engineering, feature scaling, encoding, transformation non-linéaire etc.)
- **Maîtrise de toutes les techniques d'évaluation** (les méthodes de validation-croisée, toutes les métriques d'évaluation de sklearn)
- **Maîtrise totale de Sklearn**
- **Maîtrise de Statsmodels pour les problèmes de séries temporelles.**
- Méthodes d'optimisation présentes dans Scipy
- **Analyse fréquentielle et FFT** avec scipy

Accomplissements :

- ☐ Construire un modèle de machine learning end-to-end (prétraitement, modélisation, évaluation) avec n'importe quel modèle de sklearn
- ☐ Avoir réalisé au moins un projet de série temporelle avec Statsmodels
- ☐ Savoir développer des modèles avec XGBoost
- ☐ Être capable d'expliquer clairement, pour chaque type d'algorithme, dans quel contexte il est pertinent, ses avantages, ses inconvénients, et comment l'interpréter

Ceinture Bleue - Maîtrise des Algorithmes de Deep Learning

Objectifs : Acquérir les compétences nécessaires pour concevoir et entraîner des réseaux de neurones.

- **Réseaux de Neurones :** Connaissance approfondie des réseaux de neurones Artificiels et leurs différentes architectures (MLP, CNN, RNN), leurs cas d'utilisations.
- **Frameworks de Deep Learning :** Maîtrise totale de l'un des frameworks de deep learning: TensorFlow ou PyTorch.
- **Optimisation de Modèles :** Techniques de régularisation (Dropout, L2) et optimisation (Adam, SGD).

Accomplissements :

- ☐ Avoir développé des applications de reconnaissance d'image
 - Avoir implémenté un réseau de neurones convolutionnel (CNN) personnalisé pour une tâche de classification d'images complexes, et l'avoir amélioré (augmentation de données, réglage des hyperparamètres, etc.)
 - Avoir développé au moins un algorithme de segmentation d'image
 - Avoir développé au moins un algorithme de détection d'objet (comme Yolo) en l'ayant fine-tuner.
- ☐ Avoir développé des réseaux de neurones récurrents dans des projets de modélisation de séquence:
 - Analyse de sentiment
 - Prédiction de série temporelle avec Deep Learning
- ☐ Etre capable de charger n'importe quel type de données dans Tensorflow ou Pytorch
- ☐ Etre capable de charger des modèles de Deep Learning pré-existant sur Pytorch ou Tensorflow, et les fine tuner

Ceinture Marron - Déploiement, Cloud, et Data Engineering Avancés

Objectifs : Savoir déployer des modèles dans le cloud et manipuler de grandes quantités de données.

- Maîtrise et compréhension parfaite de **Linux** obligatoire a ce stade.
- Maîtrise de **Docker**
- Compréhension et utilisation des **API**: REST, etc.
- **Cloud Computing** : Utilisation d'un des clouds provider (AWS, GCP ou Azure) pour l'entraînement et le déploiement de modèles
- **Data Engineering** : Maîtrise de SQL avancé, des outils de manipulation de grandes données (Spark, Hadoop), et des pipelines ETL.

Accomplissement :

- ☐ **Avoir développé un SaaS de Machine Learning ou de Deep Learning** sur le cloud de votre choix.

Ceinture Noire - Expert et Architecte Machine Learning

Objectifs : Devenir un expert capable de concevoir des solutions complètes et de pointe en machine learning.

- **IA Générative:** Capacité à concevoir des modèles personnalisés, expérimenter avec des architectures de pointe (Transformers, GANs, Diffusion), Fine-Tuning et utilisation de modèles provenant de HuggingFace, compréhension des recherches récentes en IA.
- **Apprentissage par renforcement:** Comprendre et savoir coder des algorithmes d'apprentissage par renforcement
- **Gestion de Projets** savoir piloter un projet de Machine Learning ou de Deep Learning de A à Z, pouvoir également identifier les red flags et les blockers d'un projet. Connaître la limite et les coûts liés aux projets de Machine Learning et de Deep Learning (Coûts d'infrastructure, comprendre les besoins en ressource d'un modèle issu par exemple de HuggingFace)
- **MLOps :** Maîtrise des pratiques de MLOps (intégration continue, déploiement continu, monitoring des modèles en production).
- **Optimisation et Scalabilité :** Compétences en optimisation de modèle et en gestion de grandes infrastructures (scalabilité horizontale/verticale, gestion de clusters).
- Maîtrise des techniques de communication et de **Storytelling**
- Connaissance dans les **régulations** IA et l'**IA éthique**, les risques et la **data gouvernance**.

Accomplissements :

- ☐ Concevoir et mettre en œuvre un projet de machine learning de bout en bout avec une infrastructure MLOps complète, avec par exemple l'utilisation de ML-Flow
- ☐ Déployer un modèle d'IA Générative sur le Cloud, comme un LLM ou un RAG.
- ☐ Innover en créant des modèles complexes adaptés à des cas d'usage avancés, avec optimisation des coûts et des performances en cloud.
- ☐ Avoir participé à plusieurs compétitions de Machine Learning (par exemple sur Kaggle) en surpassant au moins les compétences moyennes du leaderboard