

# Morgan Gautherot

## Contact

Téléphone : 0675967965

Email : gautherotmorgan0@gmail.com

LinkedIn : <https://www.linkedin.com/in/morgan-gautherot-phd-05a227108/>

You Tube : <https://www.youtube.com/channel/UCyikbI NJC ntBC l f P Oh Pb F q Q>

Organisation de conférences data :

<https://www.meetup.com/fr-FR/lille-big-data-and-machine-learning-meetup/>

## Compétences techniques

**Langages** : Python, R, SAS, HTML/CSS, Bash

**Manipulation des données** : SQL, BigQuery, Pandas, Numpy, SQLAlchemy, Spark

**Visualisation des données** : Matplotlib, Seaborn

**Machine learning** : Keras, Tensorflow, Sklearn

**Tests** : Black, flake8, Pytest

**API** : fastAPI

**Prototype** : Flask, Dash, Rshiny

**Cloud** : Google Cloud Platform, Amazon Web Services, Render

**Orchestration** : Airflow

**Gestion de code** : Git, GithubAction

**Gouvernance** : Colibra

**Conteneur** : Docker

**OS** : Linux

## Formation

Machine Learning en production (Coursera) - 2022

<https://www.coursera.org/account/accomplishments/certificate/ZW5XPYYKSFUZ>

Thèse universitaire en machine learning appliqué à l'imagerie médicale (Université de Lille) 2018 -2021

Spécialisation en IA pour la médecine (Coursera) - 2020

<https://www.coursera.org/account/accomplishments/specialization/HEJ5T5T9SRQ9>

Spécialisation en deep learning (Coursera) - 2018

<https://www.coursera.org/account/accomplishments/specialization/8WJ36KFYYGDW>

Diplôme d'ingénieur en informatique et statistique (Polytech Lille) 2012 - 2017

## Langues

Français - langue maternelle

Anglais - courant

## Expériences professionnelles

Octobre 2023 – avril 2024 | Leroy Merlin

Projet : Créer des modèles pour estimer l'efficacité énergétique d'un logement

Responsabilité occupée : Data scientist

Contexte :

Leroy Merlin est le leader français du marché de l'aménagement de la maison. La mission de l'entreprise est de rendre l'habitat plus accessible et plus durable. Dans cette optique, l'entreprise intègre à son site web un simulateur capable d'estimer la performance énergétique de chaque logement. L'objectif est de pouvoir suggérer des améliorations concrètes pour améliorer le confort de vie et contribuer à l'effort de réduction de la consommation d'énergie.

Missions :

- Analyser les données externes.
- Intégrer les données ouvertes dans les systèmes d'information.
- Création d'un modèle de prédiction de la performance énergétique d'un bâtiment (DPE, GES, déperdition de chaleur, etc.).

- Création d'une API pour mettre le modèle à disposition du front-end.
- Gestion du cycle de vie d'un modèle d'apprentissage automatique.
- Proactif sur la vision business du produit.

Environnement Technique : Google Cloud Plateforme, BigQuery, FastAPI, Python, SQL, TensorFlow, Pytest, Github

## Depuis septembre 2018 | Professeur associé

Ecole : ISEN, EDHEC, IAE, Polytech Lille

Université : Université de Lille, Dakar institute of Technology

Institut de formation : ORSYS, Jedha

Création de contenus pédagogiques et évaluation de compétences sur les thèmes suivants :

- Machine learning
- Deep learning
- Data science
- Computer vision
- Natural language processing

## Janvier 2023 – Septembre 2023 | Nomade digital

Responsabilité occupée : Data scientist, Software engineer, Enseignant

Contexte :

De janvier à juillet, j'ai exploré la région de l'Océanie, visitant la Nouvelle-Zélande, la Tasmanie, l'Australie, les Fidji et la Nouvelle-Calédonie. C'était l'occasion pour moi d'expérimenter le mode de vie de nomade digital, me permettant de travailler tout en voyageant. J'ai réussi à accomplir diverses missions de courtes durées, qu'elles soient de prestation de service, de conseil ou de formation. Ci-dessous une liste non exhaustive des missions que j'ai pu réaliser.

Projet 1 :

Mise en place du backend d'un site web destiné à l'apprentissage au moyen de cartes mémoire (flashcards)

Contexte :

Les flashcards offrent un moyen pratique et efficace d'apprentissage grâce à la révision active, la personnalisation et la répétition espacée, les rendant adaptées à une variété de sujets d'étude. Leur utilisation favorise une mémorisation plus rapide et efficace tout en rendant le processus d'apprentissage plus engageant.

Environnement Technique : Python, Flask, SQLAlchemy, Render, Github

Github : <https://github.com/MorganGautherot/Flashcards>

Livable : <https://flashcards-g5n8.onrender.com>

## Projet 2 : Création d'une formation "Deep learning par la pratique"

Contexte :

Encadrer des personnes voulant apprendre à maîtriser les algorithmes de deep learning grâce à un accompagnement individualisé et des vidéos explicatives.

Environnement Technique : Python, TensorFlow, Keras, Black, Flake8, Linux, Github

Livable : <https://www.aiforyou.fr/deep-learning-par-la-pratique/>

## Projet 3 :

Création d'une API de détection de visite.

Contexte :

Mon client recherchait une API capable de simuler l'arrivée de clients dans un magasin, dans le but de développer un POC (Proof of Concept). L'objectif était de mettre en place une API utilisable via CURL, renvoyant le nombre de visites quotidiennes.

Environnement Technique : Python, Flask, FastAPI, Pytest, Render, Github Action

Github : [https://github.com/MorganGautherot/fake\\_data](https://github.com/MorganGautherot/fake_data)

Livable : <https://fake-date-h0xo.onrender.com/>

## Projet 4 : Création d'une formation "Découvrez les structures cachées de vos données"

Contexte :

Proposer une formation gratuite à ceux qui veulent devenir data scientist afin de comprendre en profondeur les algorithmes de clustering.

Environnement Technique : Python

Livable :

<https://www.youtube.com/watch?v=soslWAHJINY&list=PLmBLgNjbSiTFvJBhVBz1MOND8pYdcrSFk>

## Projet 5 :

Réalisation de contenus de vulgarisation de la data pour grand public

Contexte :

Créer des vidéos de vulgarisation sur l'IA est essentiel pour rendre la technologie accessible au grand public, en simplifiant des concepts complexes, favorisant ainsi la compréhension générale et la préparation pour un avenir où l'IA jouera un rôle majeur.

Livrable : <https://youtu.be/7WGWcz-l4Us>

Septembre 2022 – décembre 2022 | ADEO

Projet 1 : Etudes des nouveaux jeux de données

Responsabilité occupée : Data engineer

Contexte :

Les entrepôts Leroy Merlin reçoivent chaque jour des centaines de colis. Tous ne sont pas contrôlés. Si des colis acceptés font défaut, la perte est pour Leroy Merlin. Pourtant tous les colis ne peuvent pas être contrôlés car seul une faible quantité fait défaut.

Un modèle a été créé pour déterminer si un colis doit être contrôlé. Le POC a été mis très récemment en production pour des magasins tests et doit être étendu à de nouveaux pays.

Missions :

- Collecter, nettoyer et orchestrer les pipelines de données pour les nouveaux pays,
- Examiner les données en collaboration avec les équipes métiers en vue d'améliorer leur qualité.
- Etudier les nouvelles sources de données disponibles,

Environnement Technique : Google Cloud Plateforme, BigQuery, FastAPI, Python, SQL, Kafka, Docker, Airflow, Colibra

Projet 2 : Mis-à-jour du modèle de prédiction

Responsabilité occupée : Data scientist / ML engineer / MLops

Contexte :

Les entrepôts Leroy Merlin reçoivent chaque jour des centaines de colis. Tous ne sont pas contrôlés. Si des colis acceptés font défaut, la perte est pour Leroy Merlin. Pourtant tous les colis ne peuvent pas être contrôlés car seul une faible quantité fait défaut.

Un modèle a été créé pour déterminer si un colis doit être contrôlé. Même si ROI du modèle est positif et suivi chaque semaine, le modèle n'a pas été mis à jour ni contrôlé depuis sa mise en production.

Missions :

- Monitoring des performances du modèle au cours du temps,
- Diagnostic des drifts dans les données,
- Amélioration des performances du modèle,
- Mise en production du nouveau modèle,

Environnement Technique : Google Cloud Plateforme, BigQuery, Python, SQL, Sklearn, Airflow, Colibra

Janvier 2022 – août 2022 | CHRU de Lille

Projet 1 : Modèle de prédiction de la dégradation de patient Alzheimer

Responsabilité occupée : Ingénieur de recherche en data science

Contexte :

Les patients atteints de la maladie d'Alzheimer subissent des dégradations cognitives hétérogènes et difficilement prévisibles à l'échelle individuelle.

L'objectif est de créer un modèle capable de déterminer la dégradation future d'un patient Alzheimer à partir de ces données cliniques et d'imageries.

Missions :

- Charger d'extraire les données utiles à partir des images IRM du patient,
- Analyser les données cliniques et imageries du patient,
- Trouver la variable cible (ici la pente de décroissance du MMSE qui évalue l'état cognitif du patient),
- Entraînement de modèle de machine learning pour prédiction la pente du MMSE de chaque patient (Random Forest, Gradient Boosting Tree, SVM, régression linéaire),
- Gestion de projet,
- Rédaction de documentation techniques du projet et d'un article de recherche.

Environnement Technique : Python, Sklearn, Pandas, Numpy, Linux, Git

Projet 2 : Modèle de prédiction de réponse à un médicament

Responsabilité occupée : Ingénieur de recherche en data science

Contexte :

Aujourd'hui les patients victimes de catatonie sont peu pris en charge et les traitements associés sont peu voire pas efficaces.

La benzodiazépine est un médicament censé soigner ses troubles, mais les patients ne répondent pas tous au traitement.

L'objectif est de créer un modèle capable de déterminer à l'avance si un patient est répondeur au traitement ou non à partir de ces données cliniques et d'imageries.

Missions :

- Charger d'extraire les données utiles à partir des images IRM du patient,
- Analyser les données cliniques et imageries du patient,
- Entraînement de modèle de machine learning pour prédire la réponse de chaque patient (Random Forest, Gradient Boosting Tree, SVM, régression linéaire),
- Gestion de projet,
- Rédaction de documentation techniques du projet et d'un article de recherche.

Environnement Technique : Python, Sklearn, Pandas, Numpy, Linux, Git

Projet 3 : Segmentation de l'hypothalamus

Responsabilité occupée : Ingénieur de recherche en data science

Contexte :

Pour évaluer l'évolution de maladie neurologique, les neurologues doivent pouvoir contrôler le volume de régions cérébrales au cours du temps.

Certaines régions comme l'hypothalamus sont très petites. Il est donc difficile d'obtenir une segmentation automatique de ces régions avec les méthodes classiques.

Des radiologues doivent donc segmenter manuellement les images, ce qui est une tâche longue et fastidieuse.

L'objectif est de créer un modèle d'apprentissage profond capable de segmenter automatiquement la région de l'hypothalamus.

Missions :

- Chargé de traiter et homogénéiser les images IRM,
- Analyser les données d'imageries,
- Implémentation de nombreuses fonctions de data augmentation,
- Entraînement de modèle de deep learning,
- Gestion de projet,
- Rédaction de documentation techniques du projet et d'un article de recherche.

Environnement Technique : Python, Sklearn, Pandas, Numpy, keras, Tensorflow, Linux, Git

Novembre 2018 – Décembre 2021 | General Electric

Responsabilité occupée : Ingénieur de recherche en data science

Contexte :

General Electric Healthcare, est un des trois plus grands constructeurs d'IRM au monde.

Depuis quelques années, les innovations des IRMs ne sont plus exclusivement techniques mais aussi dues à l'intégration de logiciels d'intelligence artificielle d'aide à la décision. Ces logiciels permettent d'aider les radiologues à obtenir une image de meilleure qualité, plus rapidement et d'obtenir des biomarqueurs associés à l'image.

Un de ces biomarqueurs est l'âge cérébral qui permet de quantifier la santé cérébrale d'un individu.

L'objectif était de développer un modèle de prédiction de l'âge cérébral capable de prédire la santé cérébrale à l'échelle individuelle.

Missions :

- Chargé de traiter et homogénéiser les images IRM,
- Analyser les données d'imageries,
- Implémentation de nombreuses fonctions de data augmentation,
- Entraînement de modèle de deep learning,
- Gestion de projet,
- Rédaction de documentation techniques ainsi que d'articles de recherche.

Environnement Technique : Python, Sklearn, Pandas, Numpy, keras, Tensorflow, Linux, Git



## Octobre 2017 – Octobre 2018 | Decathlon

Responsabilité occupée : Data scientist

Contexte :

Decathlon est un des leaders de la grande distribution de sport et loisirs.

L'objectif de l'équipe data science est de tirer parti de la donnée pour aider les métiers du commerce, de l'offre et de la logistique dans leurs actions de tous les jours.

Missions :

- Elaboration de modèle d'association à base de graphes,
- Prédiction de sport pratiqué grâce à la consommation des clients,
- Prévion de ventes à la semaine sur 1 an,
- Prototypages, test et mise en production des solutions.

Environnement Technique : Python, Sklearn, Pandas, Numpy, keras, tensorflow, Linux, git, dash, Rshiny, Spark, AWS

## Mars 2017 – septembre 2017 | Crédit Agricole Consumer Finance

Responsabilité occupée : Stagiaire data scientist

Contexte :

Crédit Agricole est une banque de détail qui offre une variété de services. J'étais membre d'une équipe composée de quatre personnes, travaillant dans la partie modélisation de l'entreprise pour les sujets liés au marketing et à la communication. Mon premier objectif était de développer un package d'auto-ml pour l'entraînement de modèle de score d'appétence en utilisant Spark. Le second objectif était d'utiliser ce package pour migrer des modèles de SAS à Python. Étant donné que la volumétrie de l'entreprise avait dépassé les capacités de SAS, l'intégration de Spark est devenue essentielle pour s'adapter à un environnement Big Data.

Missions :

- Veille technologique sur les nouveaux algorithmes, les techniques de machine learning et les optimisations
- Développer des solutions Big Data et des techniques d'apprentissage automatique pour le département marketing.

- Développer un programme d'auto-ML pour entraîner automatiquement des modèles de scores d'appétence

Environnement Technique : SAS, R, Python, Spark, Sklearn, Pandas, Matplotlib

## Mai 2016 – Septembre 2016 | Oney

Responsabilité occupée : Stagiaire data analyst

Contexte :

Oney est une banque de détail qui offre une variété de services. J'étais membre d'une équipe composée de cinq personnes, chacun ayant une spécialisation sur un produit spécifique. Dans le cadre de cette mission, j'étais chargé de me concentrer sur l'analyse des données liées au crédit renouvelable.

Missions :

- Analyse statistique du comportement des clients,
- Création de dashboard pour la suivit des KPI,
- Exploration du parcours client,
- Etudier l'efficacité des campagnes marketing,
- Gestion de la sur sollicitation des clients par emails.

Environnement Technique : SAS, R, Excel

## Juin 2015 – juillet 2015 | Le guide des célibataires

Responsabilité occupée : Stagiaire data scientist

Contexte :

En juin 2015, j'ai travaillé en tant que data scientist dans une startup nommée 'Le guide des célibataire'. J'ai travaillé sur leur site de rencontre qui propose des activités, des voyages et des rendez-vous aux célibataires. Durant ce stage, j'étais en charge du développement d'un algorithme de matching dans la partie rencontre de leur site. Cet algorithme permettait aux utilisateurs du site d'avoir accès à la liste des personnes qui leur convenaient le mieux.

Missions :

- Compréhension du besoin client,
- Veille technologique sur les algorithmes de matching,
- Création d'un algorithme de matching

Environnement Technique : PHP, html, JavaScript, SQL

