

# LAFORTUNE EMERICK

Data Scientist | Machine Learning | Big Data | NLP | Python | Deep Learning

+33663732677 @emerick.lafortune@gmail.com linkedin.com/in/emerick-lafortune  
github.com/elafortune/data-science-portfolio Paris

## EXPÉRIENCE

### Machine learning competition

#### Kaggle : LLM Science Exam

- 2025
- Analyse exploratoire wordcloud, analyse syntaxique, cluster classification des questions par thème
- Prétraitement tokenization des questions et réponses, conversion en vecteurs numériques, encodage
- Entraînement du modèle adapté à la quantité de données (SVM, randomforest, deep learning)
- Evaluation du modèle

### projet data science

#### Kaggle : Credit Card Fraud (2020)

- 2025
- Choix du modèle prédictif optimal parmi plusieurs modèles
- Optimisation du modèle (hyperparametre-tuning, ACP, choix des features)
- Choix de la méthode d'évaluation des métriques (kfold, fl-score/recall)
- Comparaison à un modèle plus complexe pour maximiser le recall (stacking)

### Stage de fin d'étude big data engineering

#### AirFrance KLM

- 2024
- Automatisation de la récupération et le traitement de données de vol (Pyspark et shell)
- Cycles ETL (extraction, transformation, loading)
- Gestion de données en data warehouse
- Ingestion, transformation, stockage de la donnée
- DevOps : développement, test et implémentation en production
- Gestion de projet agile (Jira, Confluence, Miro)
- CI/CD avec outils devOps tels que Jenkins, Github

### Stage software engineer (2 mois)

#### AV42

- 2022
- Data ingestion : récupération et enregistrement de données sources
- Intégration d'API pour l'automatisation de processus internes en gestion de projet et comptabilité (avec n8n, budibase, webflow)
- Gestion de projet : comptes rendus hebdomadaires, évolution suivie et optimisation en continu du livrable final (Notion)

### Projet computer vision




#### Télécom SudParis

- 2023
- Récupération d'une base de donnée de signatures
- Visualisation des signatures (seaborn)
- Apprentissage non supervisé : clustering selon des mesures de complexité (GMM et mesure d'entropie différentielle)
- Génération des catégories de signatures
- Visualisation des résultats et rapport

## RÉSUMÉ

Passionné par l'analyse de données et la science des données, je souhaite contribuer à des projets qui transforment des données complexes en solutions significatives. J'ai acquis une solide expérience en modélisation prédictive, en évaluation de modèles et une appétence forte pour les technologies python, shell et sql. J'aspire à mettre cette expérience, mon esprit d'innovation et ma rigueur technique au service de la data science.

## RÉALISATIONS

-  **Amélioration du Modèle Prédictif**  
Optimisation des métriques (recall, f1-score)
-  **Efficacité Des Processus ETL**  
Réduction du temps ETL avec Pyspark (parallélisation des calculs)
-  **Réussite Commerciale Significative**  
responsable développement commercial d'une équipe ayant généré 100k en revenus.

## CERTIFICATION

- Toeic (score :965/990)
- DeepLearning.ai: introduction to data engineering

## LANGUES

Anglais	C2
Français	C2
Espagnol	B2

## EXPÉRIENCE

### Interface de calcul du prix d'options

- 2024
- Interface de calcul du prix d'options (projet gestion des risques financiers) (python)
- Calcul du prix d'options américaines méthodes: Monte-Carlo, Résolution d'équation aux dérivées partielles par méthode des différences finies, Transformée de Fourier
- Création d'une interface graphique pour le choix de la méthode d'approximation, le choix des types d'options, et le calcul des différents grecques (tkinter)

### Echange à l'étranger (Lituanie)

#### Vilnius University

- 2023
- Programmation web full stack en MERN (MongoDB, Express, React, Node) d'une interface utilisateur de commande de livres (HTML, CSS, back-end JavaScript)
- Windows API (C#) and financial technology classes (English)

### Modélisation du prix de l'immobilier aux Etats-Unis (Data science project)

#### Télécom SudParis

- 2023
- Data visualisation (seaborn, matplotlib): modélisation des données clés
- Concatenation de bases de données (web scraping et Kaggle), nettoyage et enrichissement des bases de données (pandas)
- Création de modèles de prédiction (boosting, bagging et stacking (scikit))
- Comparaison du modèle avec un modèle économétrique (Oaxaca-Blinder)
- Évaluation de l'erreur (recall, fl-score, precision)

### Responsable développement commercial

#### Sprint Junior-Entreprise

- 2021-2023
- Création de process pour la prospection de clients (Cold Call, Mailing et Présentiel)
- Qualification du besoin client et envoi des propositions commerciales
- Signature de plusieurs missions ( + de 100k de chiffre d'affaires réalisé par l'équipe)
- Gestion de projet : missions en IA (computer vision) et développement logiciel

## ÉDUCATION

### Mathématiques appliquées et statistiques spécialisation en data et machine learning

#### Institut polytechnique de Paris: Telecom sudParis

- 2025
- Mathématiques appliquées et statistiques spécialisation en data et machine learning
- Big Data engineering data architecture, programmation (SQL, Pyspark Hadoop, Cloud), Nettoyage, manipulation et analyse de données
- Data Science, Machine Learning, Computer vision et NLP (LLM)
- Reinforcement Learning et modèles graphiques
- Financial Risk (Calcul stochastique, Produits structurés, python)
- Apprentissage des langages informatiques
- Python 6 ans d'expérience librairies (Numpy TensorFlow, scikit, Pandas, Pyspark, seaborn)
- 0 SQL: 6 ans d'expérience (Oracle, MySQL, SQLite, NoSQL: MongoDB)
- Shell (Linux): 3 ans d'expérience
- Kubernetes, Docker, Terraform: 1 an d'expérience

### Classe Préparatoire Maths Physique

#### Lycée Stanislas Cannes

- Trois années de classes préparatoires MPSI/MP (Maths sup/Maths spé)

## COMPÉTENCES

Agile LLM AWS Big Data

Computer Vision CSS

Deep Learning Docker ETL

GitHub Hadoop HTML Java

JavaScript Jenkins JIRA Keras

Kubernetes Linux MongoDB

Matplotlib Machine Learning

Hugging Face NLP NoSQL

Numpy Pandas Pyspark Python

Random Forest React Scikit-Learn

Seaborn SQL SVM TensorFlow

Terraform Tkinter Web Scraping

Webflow

## PASSIONS

Basket départementale 2



Échecs top 350 France (2270 elo)