

FLORENT LIN

Data Scientist - Machine Learning & NLP

📍 6 Av. Haroun Tazieff, 77600 Bussy-Saint-Georges · ✉ lflorentlin@gmail.com · ☎ 07 81 63 66 63
in linkedin.com/in/florentlin · 🐙 github.com/flin1

Formation Académique

ENSAE Paris - IP Paris

Cycle d'ingénieur (Probabilités, Statistiques, Machine Learning, Économétrie)

Palaiseau, IdF

Sept. 2021 - Juin 2024

- **Spécialisation:** Deep Learning, NLP, Machine Learning, Statistique Bayésienne, Séries Temporelles, Optimisation.

Lycée Saint-Louis

Classes préparatoires aux Grandes Écoles (PCSI/PC*)

Paris, IdF

Sept. 2019 - Juin 2021

Expériences Professionnelles

Consultant Data Scientist - Contrat avec SNCF Réseau

Aubay Data & AI

Boulogne-Billancourt, IdF

Depuis Janv. 2025

- Référent Data Science et ML pour l'optimisation de la maintenance prédictive des infrastructures ferroviaires.
- Ingénierie de données complexes : traitement et harmonisation de datasets multi-sources.
- Modélisation prédictive de l'apparition de défauts via algorithmes supervisés (XGBoost, Random Forest).
- Optimisation de modèles sur données déséquilibrées (SMOTE et métriques : Recall, F1-Score, ROC-AUC).
- Communication des résultats : vulgarisation, rédaction de rapports et recommandations opérationnelles.

Stage en Data Science - Machine Unlearning

Aubay

Boulogne-Billancourt, IdF

Août - Déc. 2024

- État de l'art sur le désapprentissage des LLMs (Concept Erasure, DPO, Prompt Corruption, etc.).
- Fine-tuning d'un modèle LLaMA (approche LoRA), implémentation d'un classificateur de prompt RoBERTa.
- Mesure de performances via métriques adaptées (Truth Ratio, probabilités conditionnelles, ROUGE score).

Projet semi-professionnel - Optimisation de tournées

Deloitte

Palaiseau, IdF

Nov. 2022 - Mai 2023

- Modélisation de contraintes (patients, infirmiers) et formalisation mathématique d'un problème de routing.
- Développement solutions algorithmiques pour l'optimisation de planning sous contraintes.
- Livraison d'un rapport de recommandations et d'un prototype fonctionnel (Python).

Projets Individuels

Algorithme de Matching

[global-franchises](#) 🐙


- Conception et architecture d'un système IA multi-agents (extraction de données, qualification de leads) et développement d'un algorithme hybride : filtres, scoring pondéré et recherche sémantique par embeddings.
- Technologies : Python, Supabase (PostgreSQL), Vector embeddings et LLM (API OpenAI & Google GenAI).

Synthèse de larges corpus textuels et graphiques

2025

- OCR et analyse sémantique de centaines de pages (textes, images, schémas) avec une boucle de feedback pour la fiabilisation du système d'analyse, et stockage des données sur le cloud.
- Technologies : Python, Firebase, Supabase (PostgreSQL), LLM et OCR (Mistral AI & Google GenAI).

Prédiction d'intentions d'achat

classification-shoppers 

- Feature engineering, implémentation d'algorithmes de classification et optimisation des hyperparamètres.
- Analyse et gestion du déséquilibre de classes pour prédire le comportement d'achat en ligne.

Compétences

Langues: Français (langue maternelle), Anglais (C1+), Chinois (B2)

Programmation: Python (PyTorch, JAX, Scikit-learn, Pandas), R, SQL, HTML/CSS, JavaScript (basique)

Outils Techniques: AWS, Docker, Git, Firebase, Supabase, Vercel, PowerBI, Office365, Latex