

SRIFI NAJLAA

8 rue Jean-François de La Pérouse - 91300 MASSY

☎ +33 6 05 68 58 74 ✉ najlaesrifi37@gmail.com 🌐 [srifi-najlaa](https://srifi-najlaa.github.io) 📄 [Srifi](#)

Ingénieure en data science, avec une expérience en machine learning, deep learning, traitement du langage naturel (NLP) et modèles de langage (LLMs). Passionnée par l'innovation et l'analyse de données, je recherche mon premier emploi pour appliquer mes compétences techniques à des projets ambitieux, en apportant des solutions innovantes et performantes.

Expériences

Stage - Data Scientist - Chercheur en statistiques IARD chez AXA

Avril 2024 – Septembre 2024

- Collecte et organisation des données, conception d'une architecture de base de données robuste pour l'analyse.
- Détection et résolution des problèmes de qualité des données (valeurs aberrantes, données manquantes, discrétisation des variables...).
- Développement de modèles prédictifs pour la tarification automobile et identification des facteurs critiques à l'aide de techniques de machine learning (XGBoost, Random Forest, GLM, etc.). Ces modèles ont renforcé les stratégies de tarification d'AXA en améliorant la précision des prévisions de tendances futures.

Chercheur en IA - Large Language Models (LLMs) chez VINCI

Janvier 2024 – Avril 2024

- Développement de modèles de langage (LLMs) utilisant la génération augmentée par récupération (RAG).
- Conception de modèles basés sur Mistral-7B et GPT-3.5 pour répondre aux questions issues de documents d'appel d'offres, offrant des interprétations précises des clauses concernant les systèmes d'extinction d'incendie.
- Amélioration des performances du modèle en utilisant des méthodes telles que les transformateurs de phrase pour l'embedding des requêtes et des techniques de re-ranking pour prioriser les informations pertinentes, permettant des réponses plus précises et contextuellement conscientes lors des requêtes interactives.

Stage - Data Scientist chez STMicroelectronics

Juin 2023 – Septembre 2023

- Conception et mise en œuvre d'algorithmes avancés pour améliorer la détection d'anomalies, en utilisant des techniques telles que Isolation Forest, One-Class SVM et DBSCAN, ce qui a amélioré l'identification et le contrôle des phénomènes aberrants dans les processus de fabrication.
- Exploration et évaluation du potentiel de la plateforme Dataiku, contribuant aux efforts d'innovation de l'équipe en introduisant de nouveaux outils et méthodes pour une analyse efficace des données et une prise de décision.

Formation

École Polytechnique — Institut Polytechnique de Paris

Septembre 2023– Août 2024

Master M2 Data Science

- Cours principaux : Traitement du langage naturel (NLP), Large Language Models (LLMs), Transformers, Réseau de neurones graphiques (GNN), Analyse des événements extrêmes : robustesse, valeurs aberrantes et modèles pour extrêmes, Renforcement learning, Cloud data infrastructure, Big Data Framework

École Des Mines De Saint-Étienne

Septembre 2021 – Août 2024

Diplôme d'ingénieur

- Domaines d'études principaux : Machine Learning, Deep Learning, Optimisation, Apprentissage statistique, Modèles probabilistes, Recherche opérationnelle, Computer Vision, Image and pattern recognition

Projets

Generative Adversarial Networks (GAN) pour la génération de micro-structures

Septembre 2022 – Janvier 2023

- Développement et formation d'un réseau antagoniste génératif (GAN) en Python, utilisant PyTorch, pour générer des images 2D de haute fidélité de micro-structures composites fibrées. Ce modèle a permis des simulations physiques et des calculs précis, ouvrant de nouvelles opportunités pour des applications industrielles dans la conception de matériaux avancés. (Équipe de 4)

Hi! Paris Hi!ckathon – Défi de prédiction des ventes

Décembre 2023

- Participation à une compétition de science des données où notre équipe a construit un modèle prédictif utilisant le régresseur CatBoost pour optimiser les prévisions de ventes et réduire la consommation d'énergie dans les chaînes d'approvisionnement. En incorporant des ensembles de données externes, nous avons amélioré la prise de décision, permettant aux entreprises d'aligner leur production sur la demande et de réduire la surproduction et les émissions de CO2. Notre projet a remporté la première place et le prix de l'excellence technique de TotalEnergies. (Équipe de 6)

Fast and Cheap – Optimisation des itinéraires de voyage

Janvier 2023 – Juin 2023

- Développement d'une application qui exploite les données SNCF pour mettre en œuvre un algorithme de partage de billets, identifiant des itinéraires de voyage plus efficaces et économiques par rapport à des plateformes populaires comme Trainline. Cette solution a permis des économies significatives et a optimisé l'expérience de voyage des utilisateurs. (Équipe de 3)

Compétences

Langages de programmation : Python (TensorFlow, PyTorch, Keras, LangChain, scikit-learn, Pandas,...), R, SQL, SAS, C, Matlab, Java

Outils Big Data : Elasticsearch, Hadoop, scala, Spark, Hive, Bases de données NoSQL (MongoDB, Cassandra)

Visualisation de données : Tableau, Power BI, Spotfire

Outils: Databricks, Dataiku, Akur8, GitLab

Langues : Français : Bilingue, Anglais : Courant, Espagnol : Intermédiaire