

# Rostand Surel

☎ 075-335-6106 — ✉ rostandsurel@yahoo.com — in linkedin.com/in/rostand-surel — 🌐 github.com/Manda404 — 📍 Versailles, France

## À la recherche d'un CDI — Data Scientist / ML Data Engineer — 4 ans d'expérience

“Doté d'une solide expérience en **analyse et traitement des données**, ainsi qu'un **esprit d'équipe** affirmé, je recherche activement à rejoindre une équipe d'**experts en Data Science & IA** pour mettre mes compétences au service de projets **innovants et impactants**.”

Cit. “ Si vous voulez que la **vie** vous **sourie**, apportez-lui d'abord votre bonne **humeur** .” - Baruch Spinoza (1632-1677)

## Expériences Professionnelles

### Assurances AXA

Fév. 2024 → Présent

Consultant Data Scientist, Préfecture de Nanterre, France

**Objectif** : Prédire la résiliation client à grande échelle et industrialiser les modèles pour améliorer l'efficacité marketing.

- Exploré et transformé des datasets massifs (~5M) sur **Databricks** avec **PySpark**, pour créer des variables pertinentes.
- Entraîné des modèles marketing avec **XGBoost** et **CatBoost**, pour prédire le churn via une *cross-validation robuste*.
- Structuré des pipelines de données dans des **packages Python** testés unitairement et déployés via **Azure Pipelines CI/CD**.
- Créé une API REST avec **FastAPI** permettant d'intégrer les prédictions de **churn** dans le système CRM de l'entreprise.
- Intégré une analyse de **SHAP values** dans les livrables, pour renforcer la confiance métier dans les modèles développés.

**Résultat** : +18 % de précision sur les clients à risque et amélioration des performances des campagnes de rétention.

### Safran Seats

Avr. 2024 → Nov. 2024

Stage Data Scientist, Issoudun, France

**Objectif** : Optimiser le processus de conception des sièges en simulant le confort à partir des données expérimentales.

- Nettoyé, structuré et enrichi les données de test de **confort** des sièges d'avion pour garantir leur **qualité** et **pertinence**.
- Optimisé la gestion des tests de confort des sièges d'avion grâce à un modèle de données évolutif sous **PostgreSQL**.
- Créé une application interactive avec **Streamlit** pour faciliter l'analyse des données de test et extraire des insights clés.
- Créé une **assistante IA SQL** basée sur **Large Language Model (SQLCoder)**, pour simplifier l'accès aux requêtes **SQL**.

**Résultat** : +65 % de gain de temps dans l'analyse des données de tests de confort grâce à l'automatisation et à l'interface.

### Sky TV

Oct. 2022 → Nov. 2023

Consultant Data Scientist, Milan, Italie

**Objectif** : Construire un moteur de segmentation prédictive pour cibler les clients à fort potentiel en CRM personnalisé.

- Effectué une analyse exploratoire des données avec **Spark** afin de détecter les anomalies, corriger les incohérences.
- Développé des requêtes **SQL** pour extraire les données et réalisé une **analyse exploratoire de données** avec **Python**.
- Conçu de modèles de **recommandation** basés sur l'**apprentissage profond** afin de proposer des offres personnalisées.
- Entraîné et optimisé le modèle de **XGBoost**, avec un suivi structuré des performances et des versions via **MLflow**.
- Collaboré avec l'**équipe marketing** pour exploiter les modèles de segmentation dans l'optimisation des campagnes.

**Résultat** : +24% d'efficacité des campagnes ciblées, avec une meilleure activation des segments à haut potentiel.

### Kuwait Petroleum S.p.A

Sept. 2020 → Oct. 2022

Consultant Data Engineer, Milan, Italie

**Objectif** : Moderniser l'architecture data pour fiabiliser et automatiser la gestion des données de distribution de carburants.

- Automatisé la collecte et l'ingestion des données avec **Apache NiFi**, stockées dans **HDFS** pour un traitement distribué.
- Conçu et maintenu des tables **Hive (HiveQL)** pour structurer et interroger les données de production de carburants.
- Développé des jobs de traitement distribué des données avec **PySpark**, optimisant ainsi l'efficacité des analyses.
- Orchestré les workflows avec **Airflow**, augmentant la fiabilité des processus de collecte, transformation et stockage.

**Résultat** : Pipeline de données robuste et scalable, réduisant les erreurs de collecte de **40%** sur des sources multiples.

## Applications Avancées en Data Science & GenAI

- **Incremental Boosting Framework** : Développement d'un **framework d'apprentissage incrémental** pour modèles de **boosting (XGBoost, CatBoost)**, permettant de faire évoluer des modèles en production sans **réentraînement complet**.  
**Impact** : Réduction de **30 %** du surapprentissage incrémental, meilleure **stabilité long terme** des modèles en production.
- **SignalGuard AIOps** : Développement d'une **plateforme d'observabilité** pour **microservices**, intégrant **Prometheus** (collecte de métriques), **Grafana** (monitoring temps réel) et un module **AIOps de détection automatique d'anomalies**.  
**Impact** : Automatisation de la détection d'incidents, amélioration de **50 % de la réactivité opérationnelle**.
- **BrevetAnalyzer IA** : Développement d'un **outil intelligent d'analyse brevets** basé sur une architecture **RAG (moteur de recherche sémantique, IA générative, innovation, productivité)** avec interface **Streamlit** et déploiement sur **AWS**.  
**Impact** : Recherche brevets accélérée de **65 %**, détection proactive des innovations concurrentes, adoption par R&D.
- **HealthDoc Classifier** : Conception d'un **système d'analyse automatique des rapports médicaux** grâce au **NLP (TF-IDF, Word2Vec, BERT)**, permettant de **classifier efficacement** les documents avec interface **dédiée aux médecins**.  
**Impact** : Gain de **15 % en précision**, réduction de **30 % du temps de traitement** des rapports médicaux patients.

Compétences

---

<b>Languages</b>	Python (Pandas, NumPy, Scikit-Learn), SQL	<b>MLOps</b>	MLflow, FastAPI, CI/CD, DVC, Monitoring
<b>IA Framework</b>	TensorFlow, PyTorch, LangChain	<b>Big Data &amp; Cloud</b>	PySpark, Databricks, Airflow, Nifi, Kafka, AWS
<b>IA &amp; Vector</b>	Pinecone, RAG Systems, Fine-tuning	<b>Clustering</b>	DBSCAN, K-Means, Hierarchical
<b>DevOps</b>	Docker, Git, Azure DevOps, Kubernetes	<b>NLP</b>	TF-IDF, Word2Vec, BERT, GPT, Embeddings
<b>Langues parlées</b>	Français(C2), Italien(B2), Anglais(B2)	<b>Machine Learning</b>	LightGBM, Catboost, XGBoost

Forces

---

- **Esprit analytique** et capacité à résoudre des **problèmes complexes**, avec un sens aigu de la **collaboration**.
- **Autonomie** dans la gestion de **projets**, avec une **communication claire** et une **curiosité** pour l'**innovation**.

Centres d'intérêt & activités

---

- **Loisirs:** J'apprécie **explorer** de nouveaux horizons et **découvrir** différentes cultures et endroits à travers mes **voyages**.
- **IA:** Data Science (**NLP**, **Computer-vision**, **Classification**, **Regression**, **Clustering**), Analyse et exploration des données.

Education

---

**Master** en Science des données et Intelligence Artificielle — *ECE, Lyon (France)*, **2023 – 2024**  
**Licence** en Ingénierie Informatique — *École polytechnique de Turin, Turin (Italie)*, **2017 – 2020**