

# LANCINE SÉRIFOU

## Data Engineer / Data Scientist / BI

✉ serifou.lancine10@gmail.com

☎ 06 47 97 67 42

📍 Paris, Île-de-France

🌐 www.lancine-serifou.com

## EXPÉRIENCES PROFESSIONNELLES

### MOE & Data Scientist

📅 08/2021 – Aujourd'hui

### CAB FORMATIONS

📍 Tremblay-en-France

- Recueillir l'expression des besoins métier
- Créer l'outil de planification des ateliers de formation avec Python
- Optimiser le temps et le coût du processus de la planification sous contraintes (automatisation)
- Rédiger la documentation technique
- Développer un ETL interne pour le BtoB (web scraping, API REST)
- Concevoir et entretenir la base de données BtoB et BtoC
- Analyser et assurer la fiabilité des données, interpréter les résultats
- Proposer et mettre en place des modèles de prévision robuste
- Créer des dashboards avec R Shiny et Power BI (Dataviz, Reporting)
- Effectuer la veille sur les nouvelles technologies d'analyse de données
- Assurer la formation des utilisateurs internes

Outils: Python, Pandas, Jupyter Notebook, Pyspark, R Shiny, Power BI, SQL, VBA/Excel (macros), Slack, Asana (gestion de projets)

### Ingénieur Data Scientist

📅 11/2020 – 04/2021

### ANFR

📍 Maisons-Alfort

- Recueillir l'expression des besoins et participer à la rédaction du cahier des charges
  - Développer un outil pour envoyer des alertes automatiques basées sur les données
  - Modéliser, concevoir les bases de données, créer les tables
  - Réaliser le processus d'alimentation des bases de données avec l'outil ETL Pentaho (transformation, restitution, ingestion des données)
  - Réaliser les tests et faire le choix des algorithmes de machine learning sous R avec le package Caret. Comparer les modèles de prédiction
  - Créer les applications web pour déployer les modèles de prédiction
  - Assurer la qualité de service rendu aux utilisateurs
  - Rédiger les spécifications fonctionnelles et techniques
- Outils: Pentaho, R, Rstudio, R Shiny, dplyr, purrr, Caret, R Markdown, SVN, vi, putty, python (POO), PostgreSQL, MySQL, CentOS

### Data Scientist

📅 04/2019 – 09/2019

### ANFR

📍 Maisons-Alfort

- Identifier les sources de données pertinentes pour la modélisation statistique
  - Créer un outil pour collecter automatiquement les données météorologiques
  - Réaliser le traitement et l'analyse des données collectées
  - Assurer la disponibilité et la qualité des données
  - Prédire la perturbation de la réception de la TNT en France avec les données météorologiques (séries temporelles: ARIMA et LSTM)
  - Créer les tableaux de bord pour visualiser les résultats de la prédiction
- Outils: R, RStudio, dplyr, ggplot2, plotly, R Shiny, forecast, FactomineR, MySQL, VBA/Excel (Macros, Tableaux Croisés Dynamiques, RechercheV)

## FORMATIONS

### Ingénieur en Mathématiques Appliquées

Option: Statistique/Data Science  
École d'Ingénieurs, Sup Galilée

📅 2020

📍 Université Sorbonne Paris Nord

### Master 1 Mathématiques Fondamentales

📅 2016

📍 Université Paris-Est Créteil

### Double Licence Mathématiques-Physique

📅 2015

📍 Université Paris-Est Créteil

## CONNAISSANCES

### Machine Learning/Deep learning

- Algorithmes: Random Forest, SVM, PCA, Decision trees, k-means, XGBoost, CNN, Anomaly Detection, Naive Bayes Classifier, Linear Regression, Logistic Regression, LSTM
- Bibliothèques: Caret, ggplot2, dplyr, pandas, Numpy, Scikit-learn, statsmodel, Matplotlib, Tensorflow, NLTK, Scrapy, Plotly, FastAPI, BeautifulSoup, Selenium, Seaborn, Streamlit, Prophet, Tidytext, MLLib (Pyspark)

### Langages de programmation

- R, Python, PySpark, VBA, DAX, Shell/Bash

### Bases de données

- MySQL, PostgreSQL, PL/SQL, NoSQL

### Systèmes d'exploitation

- Linux, macOS, Windows

### Logiciels ETL

- Pentaho Data Integration, Talend

### Data Visualisation & BI

- Shiny, Power BI, Pentaho, Dataiku, Streamlit, Zoho Analytics, Tableau

### Environnements Big Data & Cloud

- Spark, Azure, GCP, AWS

### Outils DevOps

- Docker, Git, SVN

## ATOUTS

Dynamique

Rigoureux

Curieux

Esprit d'équipe

Motivé

Autonome

## LANGUE

Anglais (TOEIC 785)

