

Souleymane Sall

Paris, Île-de-France 07 53 44 95 60 sallsouleymane2207@gmail.com

PORTFOLIO

MEDIUM

GITHUB

LINKEDIN

DATA SCIENTIST

COMPÉTENCES

PROGRAMMATION

- Python (NumPy, Pandas, Scikit-learn)
- R · SAS · SQL

MACHINE LEARNING

- Scoring · Segmentation
- Modélisation prédictive
- Régression, arbres de décision
- XGBoost

OUTILS & CLOUD

- GCP (Vertex AI, BigQuery)
- Dataiku
- Power BI · Excel

MLOPS

- CI/CD · MLflow · Docker
- API (FastAPI, Flask)

LANGUES

Français
Natif
Anglais
TOEIC 895 – C1
Espagnol
Notions

INTÉRÊTS

- Lecture (dev. personnel, littérature africaine)
- Sport (muscultation, football)
- Cuisine ouest-africaine

Data Scientist diplômé d'un Master en Statistique Appliquée, fort d'une expérience en entreprise et de plusieurs **projets appliqués en data science**.

J'interviens sur l'**analyse de données clients** et la construction de **modèles prédictifs** pour aider à la prise de décision.

Bonne base en pratiques **MLOps** : versioning, déploiement via API et industrialisation des pipelines.

EXPÉRIENCE

Data Scientist

Mars – Sept. 2025

MICROPOLE · GROUPE TALAN · PARIS

- Conception et déploiement d'un pipeline de segmentation client sur GCP (Vertex AI, BigQuery), automatisant l'analyse de **2 millions de clients** et identifiant **10 segments** pertinents pour le marketing.
- Construction d'une **base de données client enrichie** à partir de tables brutes, réduisant le temps de préparation des données de **60%** et facilitant l'**analyse RFM** pour l'équipe métier.
- Analyse de paniers** pour identifier les produits fréquemment achetés ensemble afin de proposer des **recommandations de ventes croisées** utilisées pour augmenter le chiffre d'affaires.
- Création de dashboards interactifs** pour suivre les performances, réduisant le temps de reporting de **40%** et facilitant l'utilisation des données par les équipes métier.

PROJETS

Scoring crédit

Mobilize Financial Services

- Modélisation du risque de défaut sur des demandes de prêt par régression logistique (AUC de **0.82**), permettant de fiabiliser l'évaluation des demandes et de réduire le taux de défaut.
- Mise en place d'une grille de score lisible par les conseillers, traduisant les résultats du modèle en critères concrets pour évaluer les demandes de crédit et aider à la décision.
- Comparaison avec SVM et XGBoost : gain de **8%** de précision avec XGBoost, documenté dans un rapport de benchmark à destination de l'équipe métier.

Détection de fraude sur transactions bancaires

Projet académique

- Analyse de transactions pour détecter des comportements frauduleux.
- Développement d'un modèle basé sur régression logistique et **Random Forest**.
- Mise en place de techniques pour mieux détecter les fraudes rares dans des données déséquilibrées.

Détection de tentatives de phishing à partir d'URLs

Mailinblack

- Analyse du contenu des URLs pour détecter des liens frauduleux.
- Identification de patterns typiques du phishing (liens trompeurs, caractères similaires, structures suspectes).
- Développement d'un modèle permettant de détecter la majorité des tentatives de phishing (**95% de précision, 93% de rappel**).
- Réduction des fausses alertes à **2%**, tout en bloquant la plupart des attaques.

FORMATION

Master Économétrie & Statistique Appliquée

2023 – 2025

Université d'Orléans

Modélisation statistique avancée · Machine Learning supervisé/non-supervisé · Deep Learning (TensorFlow) · Big Data & Cloud (SQL, GCP) · Scoring

Licence d'Économie

2019 – 2022

Université de Bourgogne

Microéconomie · Macroéconomie · Économétrie · Statistiques descriptives · Modélisation quantitative

CERTIFICATIONS

SAS Certified Statistical Business Analyst

SAS Base Programming

Python for Data Science

Dataiku Core Designer & ML Practitioner