

Saliou Diedhiou

CONTACT

- 📍 96 Boulevard Bauchi re, 83200 Toulon · Mobilit  nationale
- ☎ +33 7 68 04 81 36
- ✉ saliou.itinova@gmail.com
- 🌐 github.com/SDKisito
- 🔗 linkedin.com/in/dr-saliou-diedhiou-aa30255b
- 🔑 Permis B

STATISTIQUE & INF RENCE

- Estimation, intervalles de confiance, tests d'hypoth ses
- Puissance, taille d' chantillon, repr sentativit 
- Validation des hypoth ses, bootstrap, robustesse
- ANOVA, r gressions lin aire et logistique, GLM
- Inf rence causale : DiD, IV, Double ML

MOD LISATION

- Analyses factorielles (ACP), classification, corr lations
- Mod les de s lection et censur s (Heckman, Tobit), panel
- S ries temporelles : ARIMA, VAR, Prophet
- ML : Random Forest, XGBoost, Lasso / Ridge, clustering
- Deep Learning : PyTorch, TensorFlow

STACK

Python

NumPy

SciPy

pandas

statsmodels

scikit-learn

PyTorch

R

Stata

SQL

PySpark

AWS

Docker

Git

Power BI

Shiny

LaTeX

FORMATION

- Master DevOps / MLOps & Cloud
Paris 1 Panth on-Sorbonne
2024-2025 · 450h
- Master Data Science
Mines ParisTech – PSL
2022-2023 · 400h+
- Doctorat en  conom trie appliqu e
Universit  de Bourgogne · Dijon
2017-2022 · Bac+8

CERTIFICATIONS

PROFIL

Data Scientist Statisticien, docteur en  conom trie appliqu e et dipl m  du Master Data Science de Mines Paris. Sp cialiste de l'inf rence statistique et de la mod lisation : je choisis la m thode adapt e aux donn es disponibles, je v rifie ses hypoth ses et je mesure la robustesse des conclusions avant de les transmettre. Exp rience des analyses exploratoires, inf rentielles et multivari es, des mod les statistiques et de machine learning, et de leur restitution claire   des  quipes m tier. Enseignement des statistiques   l'universit .

EXP RIENCES

GreenOffice / Storkcom

Juin 2026 – Ao t 2026

Data Scientist & Ing nieur IA · La Rochelle (  distance) · Freelance

Contexte : Plateforme d'analyse du gaspillage alimentaire en restauration scolaire, combinant images, pes es et mod les d'IA. Objectif : produire des indicateurs fiables et choisir les mod les sur des crit res mesur s.

Missions r alis es :

- **Protocole d' valuation comparative** de 7 mod les de langage : grille de notation pond r e (qualit , hallucinations, respect du format), 350 r ponses analys es, rapports automatis s en Python
- **Aide   la d cision** : arbitrage qualit  / co t selon les volumes pr vus et recommandation argument e du mod le retenu
- **Mod lisation** : architecture de vision par ordinateur (SAM 3, RF-DETR) coupl e aux capteurs de pes e, avec boucle de r -entra nement   partir des corrections des utilisateurs
- **Contr le des sorties** : validation stricte du format, r gles de coh rence et de d tection d'anomalies sur les r sultats

Python

pandas

 valuation de mod les

Computer Vision

FastAPI

Cr dit Mutuel

F v 2026 – Avr 2026

Data Scientist / Ing nieur IA · Toulon Libert  · Freelance

Contexte : Syst me de questions-r ponses sur une documentation technique bancaire, soumis   des exigences de fiabilit  et de conformit . Objectif : mesurer et r duire le taux d'erreur en production.

Missions r alis es :

- **Indicateurs de qualit ** des r ponses et suivi statistique des erreurs en production (Prometheus, Grafana)
- **Analyse des causes d'erreur** et am lioration de la recherche documentaire (approche hybride dense + lexicale)
- **Tra abilit  compl te** des requ tes pour l'audit et la conformit  SI
- **R sultats** : 60 % d'erreurs en moins, 99,5 % de disponibilit 

Python

Indicateurs qualit 

Qdrant

Prometheus

Grafana

H pital Sainte Musse

Ao t 2025 – Oct 2025

Data Scientist · Toulon · Freelance

Contexte : Analyse de donn es patients h t rog nes (dossiers, laboratoires, imagerie) pour appuyer le diagnostic, dans un cadre RGPD strict.

Missions r alis es :

- **Structuration et contr le qualit ** de donn es m dicales multi-sources
- **Analyse de corr lations et de patterns** pour la d tection pr coce de pathologies
- **Restitution** des analyses aux  quipes m dicales

Python

Donn es de sant 

RGPD

- AWS Solutions Architect – Associate (en cours)

DataScientest – Data Science & Deep Learning

- Recurrent Neural Networks with TensorFlow (08/2023)
- TensorFlow (07/2023)
- Recurrent Neural Network with PyTorch (07/2023)
- Convolutional Neural Network with PyTorch (06/2023)
- Convolutional Neural Networks with Keras (06/2023)
- Dense Neural Networks with Keras (06/2023)
- SQL for Data Science (05/2023)
- Machine Learning with PySpark (05/2023)
- Time Series Analysis with Python (03/2023)
- Dimensionality Reduction with scikit-learn (03/2023)
- Regression with scikit-learn (03/2023)
- Clustering with scikit-learn (03/2023)
- Advanced Classification with scikit-learn (02/2023)

DataScientest – DevOps & MLOps

- Terraform (09/2025)
- Ansible (09/2025)
- GitLab (06/2025)
- Unit Testing (06/2025)
- Jenkins (05/2025)
- Kubernetes for DevOps (05/2025)
- Docker for DevOps (04/2025)
- Elasticsearch (04/2025)
- NoSQL Databases – MongoDB (03/2025)
- Relational Database (03/2025)
- Introduction à la data en entreprise (01/2025)
- FastAPI (11/2024)
- Intégration de données API (11/2024)
- Python for DevOps (11/2024)
- Bash and Linux (11/2024)

LANGUES

- Anglais : professionnel (publications)

Itinova

Oct 2022 – Jan 2026

Data Scientist · Toulon · CDI · Clients : Métropole TPM, Préfecture du Var

Contexte : Analyse de données administratives massives sur les demandes d'hébergement d'urgence et d'insertion. Objectif : comprendre les facteurs de la demande, l'anticiper et éclairer les décisions publiques.

Missions réalisées :

- **Statistiques descriptives et inférentielles** : tests d'hypothèses, régressions multiples, ANOVA, validation des hypothèses des modèles
- **Analyse multivariée** : réduction de dimension (ACP), classification, sélection de variables, identification des facteurs influents
- **Modèles prédictifs** (XGBoost, Random Forest) et **séries temporelles** (ARIMA, Prophet, LSTM) pour anticiper les demandes à fort volume
- **Détection d'anomalies** (Isolation Forest, autoencodeurs) sur les données d'exploitation
- **Restitution** à des élus et directeurs : notes de synthèse, tableaux de bord Shiny, Power BI et Streamlit ; industrialisation en Python (FastAPI, GitLab CI/CD)

Python R statsmodels scikit-learn XGBoost ARIMA PostgreSQL Power BI

Université de Bourgogne · Laboratoire d'Économie de Dijon Sep 2017 – Déc 2022
Chercheur doctorant en économétrie appliquée · Dijon

Contexte : Thèse : modélisation prédictive et causale des déterminants de l'innovation environnementale, à partir des enquêtes Community Innovation Survey (CIS) et de données Insee.

Missions réalisées :

- **Modèle structurel en 3 étapes** (Crépon-Duguet-Mairesse) : modèle de sélection de Heckman, régression censurée (Tobit), fonction de production de connaissances
- **Validation statistique** : bootstrap, tests de robustesse, vérification des hypothèses d'identification
- **Machine learning** : Random Forest et XGBoost (validation croisée stratifiée, réglage des hyperparamètres), Lasso / Ridge en grande dimension
- **Préparation de données d'enquête** à grande échelle : valeurs manquantes, valeurs aberrantes, harmonisation des variables
- **Publications internationales** en anglais ; enseignement des statistiques, de l'économétrie et de la prévision (TD et cours, Université de Bourgogne et Université de Perpignan)

R Python Stata Économétrie Inférence causale Bootstrap