

# Séphora Mitossede

Data Scientist

Data Scientist spécialisée en statistiques appliquées et Machine Learning. Expérience en développement et mise en production de modèles (ML, NLP, LLMs/GenAI), de l'exploration des données à la génération d'insights pour la prise de décision.

✉ mitossedes@gmail.com

📍 35700 Rennes

☎ +33 605732129

🌐 github.com/Sephora-MITOSSEDE

## EXPERIENCE PROFESSIONNELLE

### Alternance | Data Scientist

Institut National de la Statistique et des Études Économiques (INSEE)

📅 Sept 2025 - Août 2026

📍 Rennes, France

- Optimisation de la méthode d'imputation des revenus en évaluant plusieurs modèles (Random Forest, Gradient Boosting, MICE) par rapport à la méthode existante.
- Mise en place du modèle d'imputation optimal, permettant une réduction de 24% de l'erreur de prédiction (RMSE) par rapport à la méthode existante.
- Exploration de données (12M d'individus, plus de 600 variables) pour identifier les variables pertinentes.
- Restitution des résultats via dashboards et rapports pour la prise de décision.

### CDD | Data analyst

École des hautes études en santé publique (EHESP)

📅 Déc 2024 - Août 2025

📍 Rennes-France

- Analyse et exploitation de données en santé publique (Python, R, Excel) pour la production d'indicateurs.
- Automatisation de rapports analytiques et traitement de données d'enquêtes avec R Markdown et Sphinx.

### Stage | Statisticienne

DSPSSEL (Ministère de l'Économie)

📅 Mai 2023 - Jul 2023

📍 Cotonou-Bénin

- Conception et analyse d'enquêtes statistiques sur la mobilité urbaine, contribuant à l'étude des transports collectifs à Cotonou.
- Nettoyage, contrôle de cohérence et imputation de données sous R.

## PROJETS

- SephBot – Assistant IA RAG CV [démonstration]** : Développement d'un assistant IA permettant d'interroger mon parcours académique et professionnel à partir de documents structurés. Conception d'une architecture **RAG hybride** combinant recherche vectorielle (embeddings) et lexicale (BM25), avec reranking neuronal, garantissant des réponses fiables, contextualisées et sans hallucination. Implémentation avec **Python**, **LangChain** et **Streamlit**.
- AgriSave – Détection de maladies des plantes (CNN)** : Développement d'une application **RShiny** de détection précoce des maladies végétales à partir d'images. Intégration d'un modèle **EfficientNetB0 pré-entraîné** (transfer learning) pour la classification des pathologies. Restitution d'un diagnostic détaillé (symptômes, recommandations) via une interface interactive orientée aide à la décision.

## FORMATION

### Master Mathématiques Appliquées, Statistique parcours Data Science et IA

Université de Rennes 2 / Rennes, France

📅 Sept 2024 - Sept 2026

**Cours principaux** : Machine learning avancé - Deep learning - Big Data - Statistique en grande dimension - NLP - Optimisation pour la science des données.

### Licence 3 Mathématiques et Informatique Appliquées aux Sciences Humaines et Sociales Parcours Mathématiques-Economie

Université de Bretagne Occidentale (UBO) / Brest, France

📅 Sept 2023 - Mai 2024

### Licence 3 Statistiques parcours Statistiques Economique et Sectorielle

ENEAM / Cotonou, Bénin

📅 Sept 2020 - Oct 2021

## COMPÉTENCES

- Langages** : Python (Pandas, NumPy, PySpark), R (tidyverse), SQL (PostgreSQL, BigQuery), SAS
- Machine Learning & Deep Learning** : scikit-learn, TensorFlow, Keras, PyTorch, tidymodels
- NLP & IA générative** : LLMs, Transformers (Hugging Face), RAG, fine-tuning, embeddings, spaCy
- Data & Visualisation** : Power BI (DAX, Power Query), Tableau, Matplotlib, Seaborn, ggplot2
- Développement & Cloud** : Streamlit, Flask (API REST), RShiny, GCP
- Outils** : Git, GitHub, GitLab, Excel (VBA, macros)
- Soft skills** : esprit analytique, résolution de problèmes, travail en équipe

## LANGUES

**Français** : Langue maternelle

**Anglais** : Niveau B2