

# Ndèye Sokhna SECK

**Data Scientist**

**Expérience:** 4 ans

## Description

---

Consultante Data Scientist, je fais usage des technologies et techniques de pointe pour la collecte, le traitement et la valorisation des données sous diverses formes et volumétries. Chez Inferensia et par les missions de conseil, je conçois, développe et déploie diverses solutions aux besoins métier : analyses exploratoires de données, modélisations statistiques, dashboards, déploiement et intégration aux applications (MLOps).

## Compétences principales

- Python
- Traitement, Analyse et Valorisation de la donnée
- MLOps

## Big Data

- SQL, MySQL, Postgres, MinIO, MongoDB
- Selenium, Apache Spark, NumPy, Pandas

## ML-DL

- PyTorch, Scikit Learn, XGBoost, LLM APIs

## DevOps / MLOps

- Git, Docker, Kubernetes
- GitHub Actions, GitLab CI/CD
- MLFlow, Azure ML Studio

## Data Visualisation

- Grafana, Plotly, Dash, Matplotlib, Solara, Streamlit

## Langages

- Python, Java, Scala

## Cloud

- Microsoft Azure

## Gestion de projet

- Déclinaison des feuilles de route
- Conception des architectures SI/Data
- Méthode Agile (Scrum)

## Langues

- Français, Anglais

## Formation

---

**Polytech Sophia Antipolis**

Diplôme d'ingénieur en Informatique

Spécialisation : Data Science

## Expériences & Projets

---

### Inferensia : Juillet 2022 – Aujourd'hui

**Contexte** Inferensia innove dans la gestion du Risk & Compliance et déploie sa plateforme de détection de fraude comptable de même que divers modules tels que la conformité de modèles IA aux lois en vigueur. En ce sens, les travaux effectués en interne s'articulent autour du développement et de l'optimisation des algorithmes de détection de fraude, la recherche et la déclinaison d'approches performantes par l'IA, la conception et la mise en production de solutions innovantes de même que leur intégration à la plateforme.

#### Missions

- Optimisation de la performance des modules de détection de fraude par leur traitement avec Spark.
- POCs et expérimentations d'approches par l'IA pour une valorisation des expériences utilisateur antérieures.
- Développement et intégration de modules de conformité à la plateforme : conception fullstack.

**Technologies** Python, MLOps (MLFlow, Airflow) / DevOps (Docker, Kubernetes), Spark, MinIO, Angular, Flask

### Relyens : Juillet 2022 - Août 2023

**Contexte** Dans le contexte d'une mission pour le compte d'Inferensia au sein d'une entreprise innovante en pleine croissance, l'objectif principal consiste à explorer et valoriser les données pour divers cas d'usage, en travaillant sur leur collecte, traitement et analyse de même que la conception d'outils pour les besoins exprimés.

#### Missions

- Collecte de données sous différents formats depuis plusieurs démembrements et équipes, l'open source data ou stockages internes.
- Data mining, Data cleaning, Data quality assessment.
- Outils développés :
  - Cartographie de données selon différentes variables sur plusieurs années.
  - Outil de recherche et de "question answering" sur des documents scannés - Fine-tuning des modèles utilisés.
  - Intégration de fonctionnalités de classification / prédiction avec du Machine Learning.
  - POCs Streamlit ou notebook Jupyter sur divers sujets.

**Technologies** Python, Machine Learning, Deep Learning, NumPy, Pandas, Dash, Selenium, BeautifulSoup

### Cetim : Septembre 2020 - Juin 2022

**Contexte** Dans le cadre du traitement, de la visualisation et de l'analyse de données IoT, l'équipe est chargée de définir et mettre en place l'architecture système d'information (SI) adaptée. Le processus comprend la récolte, le traitement et la sauvegarde des données au format h5 avec Pytables, leur classification ainsi que la génération de fichiers HTML pour la visualisation des prédictions à l'aide de Plotly. L'objectif principal est de définir les besoins SI et Data pour la mise en place du projet à grande échelle dans l'environnement Azure, d'analyser les données pour permettre la maintenance prévisionnelle grâce à des techniques de Machine Learning.

#### Missions

- Définition des besoins SI/Data selon les besoins et objectifs exprimés.
- Design de l'architecture SI (Edge/Cloud Azure) – environnements de stockage, de développement (MLOps/DevOps) et de production (Conteneurisation) dans Azure.
- Déclinaison des cahiers de charge et de la feuille de route.
- Suivi de l'implémentation des services.
- Reporting.

**Technologies** Python, Pytables, Plotly, Azure (Edge/Cloud), MLOps/DevOps, Conteneurisation.

### Amadeus : Janvier 2019 - Juin 2020 - Alternance

**Contexte** La conception d'une application de détection d'anomalies lors de la génération de fichiers d'inventaire de vols est au cœur des préoccupations. Cela implique l'extraction et le traitement des logs, la déduction et la visualisation des KPIs via une interface utilisateur graphique (GUI), ainsi que l'automatisation du pipeline. De plus, il est nécessaire de concevoir et d'entraîner un modèle de détection automatique d'anomalies utilisant l'apprentissage automatique avec Scikit Learn. Enfin, une conception fullstack de l'application web est prévue.

**Missions**

- Extraction et traitement des logs.
- Déduction et visualisation des KPIs à travers une interface utilisateur graphique (GUI).
- Automatisation du pipeline de traitement des données.
- Conception et entraînement du modèle de détection automatique d'anomalies en utilisant l'apprentissage automatique avec Scikit Learn.
- Conception fullstack de l'application web.

**Technologies** Scikit Learn, Python, C++, Fullstack Web Development.

## CNRS : Avril 2018 - Octobre 2018 - Stage

**Contexte** Dans le cadre de l'édition d'une plateforme destinée à la modélisation et l'édition d'un réseau de gènes, il est nécessaire de procéder au refactoring et à l'intégration des outils. Cela implique la lecture et la traduction des fichiers de description du réseau en objets, ainsi que leur intégration avec les logiciels Java de la plateforme.

**Missions**

- Lecture et traduction des fichiers de description du réseau en objets.
- Intégration avec les logiciels Java de la plateforme.

**Technologies** Java, Bio-informatique.