

Aina KIKI-SAGBE

Tél. : +33 6 05 91 36 85

Courriel : aina.kiki.sagbe@gmail.com

Résidence : France (Ile-de-France)

Langue : Français

(Lieux de Mission : **France** « métropole », **Canada** « Québec, Sherbrooke, Montréal »)

## Data Engineer / Machine Learning Engineer / Data Scientist

### PROFIL :

<https://github.com/AinaKIKISAGBE/profil>

Expérimenté en data science, data engineering et machine learning, j'industrialise des pipelines de données (Python, PySpark, SAS, SQL, ...), en environnement Big Data (AWS, GCP, DATAIKU). Je suis force de proposition et j'ai l'habitude de travailler en équipes multidisciplinaires et de mener des projets de bout en bout. J'occupe également le rôle de chef de projet techniques ce qui implique du management d'équipe. J'ai formé également des professionnels en IT et Data en entreprise ainsi que sur ma plateforme itech14.com.

### COMPÉTENCES CLÉS

- Data & Big Data : Python, R, SAS, PySpark, SQL, Hadoop, Spark, BigQuery, Dataiku, CI/CD, Airflow
- Cloud : Google Cloud (BigQuery, Dataproc, Vertex AI...), AWS (S3, lambda, EC2...), Serverless
- Machine Learning : Classification supervisée et non supervisée, Deep learning, tensorflow
- DevOps : Terraform, Ansible, Docker, Gitlab CI, CI/CD, Kubernetes, API REST (Flask, Fastapi), Shell linux
- Bases de données : Oracle, PostgreSQL, Hive, Cosmos DB, MySql
- Visualisation & Reporting : Looker Studio

### DIPLOMES

- **Fr : Master 2 / Ca : Maîtrise (2e cycle)** Mathématiques Appliquées et Statistiques | **France - 2018**
- **Fr : Licence 3 / Ca : Baccalauréat (1er cycle)** Mathématiques, Informatique et Économie | **France - 2016**

### CERTIFICATIONS:

- Dataiku (Full certifié): [[Advanced](#), [Core](#), [Dev](#), [ML](#), [MLOps](#)] academy.dataiku.com (2023-2024)
- Machine Learning Engineer (MLOps) [[afficher le certificat](#)] blent.ai (2022)

Lien vers les certifications: <https://github.com/AinaKIKISAGBE/profil>

### EXPÉRIENCES PROFESSIONNELLES

#### Formateur PYSPARK ET AZURE

08/2025 – En cours

ALTEN

| FRANCE

- PYTHON : Initiation
- PYSPARK : Initiation
- Maîtriser les concepts fondamentaux d'Apache Spark et PySpark
- Apprendre à manipuler et transformer des données à grande échelle
- Comprendre l'architecture distribuée de Spark
- Savoir optimiser les performances des applications Spark
- Intégrer PySpark avec les services cloud (Azure)
- Différents mode de connexion selon si le conteneurs est (HNS) ou (non HNS)
- CLOUD AZURE blob : Initiation
- Création de conteneurs (HNS) et (non HNS)
- Création de clé et autorisation

**Environnement** : Python, PySpark, SQL, Unix Shell, Cloud AZURE

### Data Engineer & ML

04/2023 – 07/2025

ALTEN → BANQUE POSTALE ASSURANCE (LBP puis CNP)

| FRANCE

- DATA ENGINEER :
  - Développement et gestion d'architectures de données pour supporter des analyses et des modèles.
  - Référent Dataiku : création de projets, automatisation via scénarios.
  - Construction et maintenance de pipelines de données (Python, PySpark), pour le stockage.
  - Automatisation de scripts (SAS, batch, VBS) et migration vers Unix (Shell).
  - Rédaction de documentation fonctionnelle, support et formation aux équipes internes.
  - Chef technique de projet et Management: définition, planification, attribution, suivi, réévaluation et priorité des tâches.
- MACHINE LEARNING :
  - Scoring : EDA, traitement des valeurs manquantes et feature selection, Machine learning (Xgboost), Evaluatuaion des models et back testing
  - Technique de Classification sur variable déséquilibrée

**Environnement** : SAS, Python, PySpark, Dataiku, SQL, Hive, Unix Shell, Batch, VBS.

### Machine Learning Engineer (MLOps)

01/2023 – 04/2023

MARKETMATER

| FRANCE

- Optimisation du taux de conversion utilisateur via des modèles de recommandation
- Développement de pipelines d'industrialisation (MLE) et de modèles de prédiction (Machine Learning).
- Déploiement de modèles via API REST sécurisées (Flask), intégration CI/CD, conteneurisation Docker.
- OCR et Deep Learning pour la lecture automatisée de pièces d'identité et documents.
- Création de tableaux de bord Looker Studio pour le suivi des indicateurs de performance (ventes, campagnes marketing, etc.).

**Environnement** : Python, GCP (BigQuery, Airflow), Docker, CI/CD, PostgreSQL, Django, MLFlow, Vagrant.

### Data Scientist ML

01/2019 – 09/2022

KANTAR

| FRANCE

- Migration de programmes (de SAS vers WPS, puis vers Python et R).
- Réentraînement et optimisation des model statistiques pendant la migration (Machine learning )
- Text mining (NLP), analyse et recodage automatique de texte.
- Formation des équipes internes sur des méthodes avancées de Machine Learning (Deep Learning, DBSCAN).
- Chef technique de projet et Management : planification, attribution, suivi, et priorité des tâches.

**Environnement** : SAS, WPS, Python, R, SQL.

### Ingénieur Machine Learning

03/2018 – 09/2018

EDF LAB R&D

| FRANCE

- Classification automatique de signaux dans le secteur énergétique.
- Extraction et transformation de données depuis Oracle (SQL, R).
- Clustering (Classification non supervisée) et détection d'anomalies
- Classification (supervisée) des signaux : bases train et test, EDA, ML avec Cross Validation
- Conception mathématique et développement informatique d'un Algorithme de Machine learning pensé spécifiquement pour le sujet « sphère de vibration du noyau »
- Développement de scripts R pour la mise en production des modèles (Knn, Forêts aléatoires, SVM, etc.).
- Minimisation des erreurs et sélection du meilleur modèle
- Analyse des probabilités affectation et évaluation de la qualité des séries classifiées
- Rédaction de rapports, présentation des résultats et mise en place de bonnes pratiques de code (Git).

**Environnement** : R, SQL, Oracle, Gitlab.

**Distinction**: Proposition de dépôt de brevet suite aux résultats

### Data Scientist / Machine Learning

05/2017 – 08/2017

IRI-FRANCE

| FRANCE

- Conception et intégration d'un programme pour l'estimation automatique (Machine Learning).

- Automatisation de processus (Unix Shell, Windows Batch).
- Rédaction de rapports et présentation des résultats aux équipes.
- Technique de d'inférence sur variable quantitative

**Environnement :** R, SAS, PL-SQL, Unix Shell, .bat .

**Distinction:** prime exceptionnelle suite aux résultats

## PROJETS PERSONNELS

---

### Formateur IT & DATA

08/2023 – En cours

itech14.com

| FRANCE

Formations réalisées : **<https://itech14.com/courses/>**

- f1 Concevoir et Déployer une flask api puis l'envoyer sur Github
- f2 Docker, GitHub & Déploiement Flask
- F3\_ci\_cd\_1\_Heroku : Maîtrise la CI/CD moderne avec Heroku
- F3\_ci\_cd\_2\_AWS : CI/CD sur AWS, DevOps (ECS, EKS, Lambda, Terraform, Ansible)
- F4\_DATA\_ENGINEER (En cours) : Data Engineering GCP & AWS : ETL moderne, FastAPI Sécurisée

Partie 1 : Devenir Data Engineer – Préparer son Environnement et le Pipeline

- Construire un pipeline Data moderne : formation complète Data Engineer
- Préparer son environnement Data Engineering pour un projet Bitcoin

Partie 2 : Extraction de Données Bitcoin – API, Pagination, Stockage Cloud

- Extraction intelligente de données Bitcoin : API, pagination & .env (Python)
- Extraction flexible : 7 jours de données Bitcoin, robuste & adaptable (Python)
- Enregistrer et envoyer ses données brut Bitcoin sur AWS S3

Partie 3 : ETL Cloud – Transformation, Chargement & BigQuery

- Pipeline ETL : Transformation & chargement de données Bitcoin sur BigQuery (GCP)
- gcp credential : Créer un projet GCP, compte de service et clé JSON
- Créer pipeline Data Engineering : ETL sous GCP avec Apache Beam
- Exporter échantillon de données pour calibrer la fonction de transformation Beam
- Fonction de transformation & Load dataset to gcp

Partie 4 : Exposer, Sécuriser et Authentifier son API

- Exposez FastAPI et BigQuery et le sécuriser contre les injection sql
- Test de sécurité injection sql api sécurisée vs api non sécurisée
- Ajouter l'authentification par token (JWT) à votre API FastAPI
- Générer une clé API JWT depuis une interface web Python
- Créer un système complet d'authentification et génération de clés API JWT

Partie 5 : Sécurité Avancée – Sniffing, HTTPS et Gestion des Clés

- Test de vulnérabilité sniffing puis sécurisation de l'api
- Données interceptées : Pourquoi chiffrer votre API en HTTPS ?
- Sécurisation contre vulnérabilité sniffing de l'api en local avec mkcert
- Sécuriser son API pour le web : audit & bonnes pratiques avant la mise en production
- Déployer une API sécurisée avec Nginx, HTTPS et gestion des clés
- Architecture finale : FastAPI, Nginx & HTTPS sécurisé
- Tester l'inscription, la génération et l'utilisation d'une API Key sécurisé

Partie 6 : Supervision & Monitoring du Pipeline de Données

- (Prometheus) Sécuriser et monitorer une API avec Prometheus & Grafana
- (Grafana) Créer des dashboards interactifs avec Grafana pour monitorer son API
- (Alert Grafana ) créer des alertes personnalisées

Partie 7 : Orchestration et Automatisation avancée avec Airflow

- Installer Apache Airflow avec Docker pour orchestrer ses ETL
- Connection ssh entre machine host distant et machine conteneur docker airflow
- dag execution: Orchestration ETL complète : déclencher un pipeline avec Airflow

Partie 8 : A suivre ...