



Frimpong Michel-Ange ADOTRI

EN RECHERCHE D'UN CDI EN TANT QUE
DATA ENGINEER

CONTACT

Noisy-Le-Grand, 93160, FRANCE

frimpong.adotri@gmail.com

[Frimpong.A.](#)

[notorious-mca](#)

COMPÉTENCES

- ▶ Langages : PYTHON, Java, Go, Scala, SQL, Bash
- ▶ CI/CD : GitHub Actions, Jenkins
- ▶ Gestion de projet: Jira, Confluence
- ▶ Big Data: APACHE SPARK, Apache Kafka, Hadoop(Hive, HDFS)
- ▶ Bases de données: MongoDB, PostgreSQL, Cassandra
- ▶ Conteneurisation : DOCKER, Docker-compose, Docker-Swarm
- ▶ Projet Versioning : GIT, GitHub, BitBucket
- ▶ Cloud: Cloudera, Azure (Azure ML Studio, Azure Data Factory), Google Cloud Platform

SOFT SKILLS

- ▶ Curieux
- ▶ Proactif
- ▶ À l'aise à l'oral
- ▶ Rigoureux

LOISIRS

- ▶ Football
- ▶ callisthénie
- ▶ Cuisine
- ▶ Gaming

LANGUES

- FRANÇAIS : lu, écrit et parlé
- ANGLAIS : lu, écrit et parlé

Aurolé de deux ans d'expériences, dont un an en tant que Data Scientist et une autre année en tant que Data Engineer, je suis titulaire d'un diplôme ingénieur en Big Data & Machine Learning à EFREI-Paris. Je perçois le métier de Data Engineer comme une expérience similaire à un jeu de LEGO où les briques d'une pipeline s'imbriquent pour former un ensemble cohérent. Ainsi, j'est passionné par ma première expérience à Informatique-CDC que je souhaite poursuivre mon parcours professionnel en tant que Data Engineer.

EXPÉRIENCE PROFESSIONNELLE

CDC INFORMATIQUE | **Data Engineer**
CDC-Informatique | Octobre 2023 - Septembre 2024

En tant que Data Engineer à la Direction des Études Informatiques, je travaille principalement sur un projet dont la finalité est de produire un reporting sur les risques de crédits, de concentration sectorielle et géographique. Je fais parti d'une équipe de MOE dont 3 data engineers, 1 Tech Lead et 1 chef de projet. Nous travaillons en mode Agile, Scrum avec des sprints de 3 semaines en moyenne, avec une modélisation Data Vault des données. Mes principales tâches sont:

- L'ingestion de tables de paramétrage
- La collecte de données
- L'alimentation du datamart
- La consolidation des données dans le Data Hub

Langage/Technologies: Python, Spark, Hadoop(Hive, HDFS), Jira, Confluence, Bitbucket, Cloudera



Stage / projet transverse en NLP
EFREI-Paris & Société Générale | Novembre 2023 - Février 2024

Les analystes de risques de marché doivent expliquer quotidiennement les variations de marché remontées sous forme d'alertes. Afin de leur alléger la tâche, les équipes de la Société Générale ont développé un algorithme qui génère des commentaires automatiques. Cependant, ces derniers sont souvent imprécis. À la tête d'une équipe de stagiaire, mon équipe et moi avions pour principale mission de dégager des axes et des méthodes d'amélioration de l'algorithme par :

- La construction d'un dataset pour la détection d'entités dans les commentaires : à l'aide du NER
- L'identification des divergences entre commentaires utilisateur & algorithme : à l'aide de NLP Clustering par ex ou autres méthodes. Cela permettrait de regrouper les couples de commentaires par similitudes, attributs
- L'extraction des spécificités de chaque cluster afin d'apporter des points d'amélioration sur l'algorithme existant

Langage/Technologies: Python, Deep Learning(Encodeurs), NLP(Bert, Knowledge Graphs), GitHub, Artifactory



Data Scientist
MUTEX | Septembre 2022 - Octobre 2023

En tant que Data Scientist au sein de l'équipe Data&Analytics de Mutex, j'ai travaillé sur le projet d'étude prédictive sur l'absentéisme au sein de la CCN Sport au sein d'une équipe de 1 actuaire et 2 Data Scientist. J'ai eu un rôle de support des actuaires en entraînant et en améliorant un modèle de clustering (DBSCAN) afin de confirmer leurs analyses. J'ai principalement réalisé:

- Le développement de modèles prédictifs pour la prévoyance collectives
- Le traitement et nettoyage de données
- La data visualization

FORMATIONS

Cycle ingénieur - parcours Big Data and Machine Learning | 2022 - 2024

EFREI-Paris, Villejuif 94800
Machine Learning, Data Visualization, Data Engineering, Cloud Computing, DevOps, Functional programming with Scala, ...

Double licence Mathématiques-Informatique | 2019 - 2023

Université Paris Cité, Paris 75006
Mathématiques pour la Modélisation & Simulation (Analyse pour l'ingénieur, Algèbre, Transformées de Fourier, Statistiques inférentielles)
Informatique (Systèmes d'exploitation, Réseau, Algorithmique, POO, Théorie des langages)

PROJETS

Build Data Platform (voir [ici](#))
Projet tutoré | Juin 2024 - Août 2024

- Mettre en place une data plateforme dans laquelle les données seront stockées dans le cloud sous format Iceberg avec un catalogue Rest. La plateforme sera orchestrée et monitorée.

Langage/Outils : Google Cloud Storage, DuckDB, Dagster, Docker, Python, Apache Iceberg

Big Data Application Project (voir [ici](#))
EFREI-Paris | Janvier 2024 - Février 2024

- Mise en place d'une infrastructure mettant en exergue un modèle de machine learning en production (API), un système de monitoring avec Prometheus et Grafana

Langage/Outils : Python, Grafana, Prometheus, Microsoft Azure, Docker

Data Engineering Project (voir [ici](#))
EFREI-Paris | Mars 2023 - Mai 2023

- Mise en place de l'architecture Docker (Docker Compose)
- Initialisation et configuration d'une stream Kafka
- Envoi des données de la stream Kafka vers S3
- Développement de 2 microservices de Data Visualization et de monitoring et ajout des microservices à l'architecture Docker via Docker Compose

Langage/Outils : Scala, AWS S3, Docker, Python, Streamlit, Apache Kafka, Apache Spark, GitHub Actions