

Aina KIKI-SAGBE
 Tel : +33 6 05 91 36 85
 Mail: aina.kiki.sagbe@gmail.com
 Pays : France
 Région parisienne : Île-de-France

Data Scientist / Data Engineer Machine Learning Engineer

Python, R, SAS, PySpark, SQL

En résumé : Formateur en Machine Learning, Développeur, Dynamisme, Travail collaboratif, Autonome

Compétences techniques

Cloud Services GCP :	Bigquery, Dataproc, Cloud Storage, cloud composer, Dataflow, Vertex AI, Data Studio, Airflow, Looker Google Studio
Cloud Services Azure :	Data factory, databricks, blob storage, data lake, cosmos db, sql db, ETL, ELT
AWS :	S3, RDS, EC2
DATAIKU	DSS, Python, Pyspark, Hive, SQL
DevOps :	Vagrant, Docker, Gitlab CI, VM, Serverless, Docker-Swarm, CI/CD
Big Data Tools :	Apach spark, spark SQL, hadoop, PySpark, BigQuery, kedro, AutoML, Tests logiciels, Tests modèles, MLflow, API REST, Flask, serverless, Docker, pipelines CI/CD, Kubernetes, MLOps, Apache Airflow, Identité et sécurité d'API
Machine learning :	Classification non supervisée : Hiérarchique kmeans, CHA « Hclust », AGNES et DIANA », DBSCAN, Clustering flou (fuzzy) « Fanny » Classification supervisée : KNN, Forêts aléatoires, Extra-Trees , Gradient-boosting, deep learning, apprentissage par renforcement (Q-learning), regression logistique Prévision de séries temporelles : ARIMA, LSTM Regression: regression lineaire, svm_lineaire, réseaux de neurones, Ridge, Lasso, Elastic Net
Langages :	SQL, Python, R, PySpark, SAS, VBA-Excel
Shell System:	Unix-Shell-bash, Windows-CMD-Batch
SGBD relationnels :	Oracle, PostgreSql
Développement Web:	Django, HTML, CSS, Javascript, liquid (Shopify)
Visualisation de données:	Rshiny, Looker Google Studio
Bureautique :	Word; Excel; Publisher; Access; PowerPoint, OneNote
Autres Outils :	SPSS, STATA, WinRats, E-views, EPI INFO

Diplômes

2016-2018	Master 2 Mathématiques Appliquées et Statistiques Université de Rennes 1	Rennes, France
2015-2016	Licence 3 Mathématiques, informatique et économie Université de Rennes 1	Rennes, France

Formations & Certifications

2023-2024	DATAIKU : Full Certification [certif_1 , certif_2 , certif_3 , certif_4 , certif_5]	academy.dataiku.com
2023	GCP : Certification Professional Data Engineer (en cours)	cloud.google.com
2023	Certification Azure Data Engineer Associate de Microsoft (DP-203) (en cours)	Coursera.org
2022	Machine Learning Engineer (MLOps) [afficher le certificat]	blent.ai
2022	Certification « Machine Learning Engineer » [afficher le certificat]	Udemy.com
2022	Certification « Spark avec Python - Pratique avec le Big Data » [lien]	Udemy.com
2020	Installer un serveur de messagerie sur Debian9 avec PostFix.	Udemy.com
2020	Certification « Django : Développer un Site de E-Commerce en Python » [lien]	Udemy.com

BANQUE POSTALE ASSURANCE – Data Engineer

Début : 04/2023 – en cours

Projet : Création de datamart, automatisation batch et vbs, migrations des programmes vers un nouvel environnement Unix, accompagnement de l'équipe sur le changement d'environnement de windows vers Unix.

Mission :

- Proposition d'architecture de données
- Référent Dataiku : création des projets sur Dataiku et automatisation des scénarios.
- Création de packages maison (customisés en SAS) pour le département et formation
- Automatisation et paramétrage des nœuds de lancement des process
- Optimisation des temps de traitement machine et explication du code
- Documentation fonctionnelle et mode d'emploi des process
- Migrations des programmes vers un nouvel environnement Unix
- Accompagnement de l'équipe sur la prise en mains de l'environnement Unix, explication et création de mini KPI pour faciliter la prise en mains.

Environnement technique : SAS, Python, Pyspark, Dataiku (DSS), SQL, Hive, Unix Shell, Batch, vbs

MARKETMATER - Machine Learning Engineer (Mlops) / IA

01/2023-04/2023

Projet : L'objectif est d'optimiser le rendement des sessions utilisateurs sur le site internet en optimisant les propositions d'articles susceptibles d'intéresser l'utilisateur suivant son comportement session.

Mission :

- Récupération et transformation des données en streaming
- Machine Learning (prédiction)
- Pipeline d'industrialisation de Machine Learning (MLE)
- Versionning et déploiement en API REST sécurisée avec Flask
- MLOps: streaming et mise à jour automatique des modèles de ML
- Autres travaux :
 - OCR et Deep Learning pour la lecture automatique de pièce d'identité et de justificatif de domicile et Ciblage marketing et emailing
 - Dashboard avec Looker Studio : pour suivre le chiffre d'affaire et d'autres indicateurs de performance d'activité ; pour évaluer l'efficacité des campagnes marketing ; ...

Environnement technique : Python, SQL, GCP, Looker Studio, Unix-Shell-bash, Docker, CI/CD, API REST (FLASK), Github, PostgreSQL, Django, HTML, CSS, Javascript, MLFlow, AirFlow, Vagrant, Agilité Scrum

KANTAR - Machine Learning Engineer / Data Scientist

01/2019 – 09/2022

Projet PRISM : projet de migration de logiciel (de SAS vers WPS, puis vers R sur le long terme). Ce projet a mobilisé plusieurs divisions dont la division en charge de fournir des programmes statistiques qui étaient ensuite tournés en tâche de fond quand l'utilisateur sélectionne ses options dans l'interface utilisateur.

Mission : Migration de SAS vers WPS, Python et R

- Réalisation de SPEC puis traduction en programme WPS et SAS
- Migration des programmes de SAS vers WPS, Python et R

- Réentraînement du model sur R puis exportation ses résultats vers un format utilisable par WPS (car certaines tables de model logistique de SAS ne pouvaient pas être utilisées correctement par WPS),
- Coding en python et faisait exécuter le programme python depuis les scripts WPS qui envoyait les paramètres en «.json ». Ensuite quand on trouvait de solution à ces scripts sous WPS, on faisait la mise à jour en WPS.
- Atelier de formation aux méthodes de Machine learning (Deep learning, DBSCAN)
- Text mining (NLP), analyse et recodage automatique de texte. L'analyse et le WorkFlow d'entraînement a été réalisé puis présenté avant la mise en production

Environnement technique : SAS, WPS, Python, R

EDF LAB R&D - Ingénieur Machine Learning / Data Scientist

03/2018 – 09/2018

Projet : Classification automatique de séries temporelles. Développement de méthode de traitement du signal dans le secteur énergétique puis sa classification. Mise en réseau du suivi projet sur le GIT d'EDF.

Mission :

- Analyse de données (base Oracle avec du SQL dans R)
- Importation d'une partie de la base Oracle dans R
- Transformation et extraction de caractéristiques sur les séries temporelles
- Caractérisation de chaque type de séries temporelles
- Clustering (Classification non supervisée) et détection d'anomalies
- Classification (supervisée) des séries temporelles :
 - Préparation de la base épurée : base d'apprentissage et de test
 - Sélection de variables (par la méthode backward dans R)
 - Validation croisée sur la base d'apprentissage
 - Prédiction des séries temporelles (variable qualitative multinomiale)
- Algorithme customisé de Machine learning « sphère de vibration du noyau »
- Régression logistique multinomiale, Knn (k-plus proches voisins), Forêts aléatoires, Extra-Trees : Extremely randomized Trees, Gradient-boosting, Svm-radial gaussien
- Minimisation des erreurs et sélection du meilleur modèle
- Analyse des probabilités affectation et évaluation de la qualité des séries classifiées
- Rédaction de script R (modifiable pour faciliter la mise en production) de l'automatisation de la classification et de l'évaluation de la qualité des séries
- Rédaction de rapports et présentation des résultats

Environnement technique : R, SQL, Oracle, Gitlab

IRI-FRANCE Chambourcy - Data Scientist / Machine Learning

05/2017 – 08/2017

Projet : Conception et intégration d'un programme d'intelligence artificielle (machine learning) réalisant des estimations.

Mission :

- Prise en main des process et environnement puis Machine Learning pour l'estimation
- Création d'un programme de lancement avec demande de profil et mot de passe
- Rédaction de rapports et présentation

Environnement technique : R, SAS, PL-SQL, Unix-Shell-bash, Windows-CMD-Batch

Loisir : Jeux de dame, foot de salle, Escape Game