Aina KIKI-SAGBE

Mail: aina.kiki.sagbe@gmail.com

Data Scientist

Data Engineer / Machine Learning Engineer

Python, R, SAS, Spark, SQL

En résumé : Dynamisme, Relationnel, Partage intellectuel

Compétences techniques

Cloud Services Azure :	Data factory, databricks, blob storage, data lake, cosmos db, sql db, ETL, ELT
Cloud Services GCP :	Bigquery, Dataproc, Cloud Storage, cloud composer
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, Svm, deep learning, apprentissage par renforcement (Q-learning), regression logistique, AdaBoost, ANN Prévision de séries temporelles : ARIMA, LSTM Regression: regression lineaire, svm_lineaire, réseau 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
Visualisation de données:	Rshiny
Bureautique :	Word; Excel; Publisher; Access; PowerPoint, OneNote
Autres Outils :	SPSS, STATA, WinRats, E-views, EPI INFO

Diplômes

	Master 2 Mathématiques Appliquées et Statistiques	
2016-2018	Université de Rennes 1	Rennes, France
2015-2016	**Licence 3 Mathématiques, informatique et économie** Université de Rennes 1	Rennes, France
2012-2013	**Licence 3 Statistiques** ENEAM (École Nationale d'Économie Appliquée et de Management)	Cotonou, Bénin

Formations & Certifications

2022	*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

Projet de certification
Machine Learning Engineer (MLOps)
05/2021 – 10/2022

Projet : L'objectif est d'étudier les événements utilisateurs sur une plateforme E-Commerce et d'optimiser les offres ciblées d'opérations marketing en proposant des réductions pour les utilisateurs pendant leur parcours d'achat. Pour cela, on s'intéresse à savoir si, au cours d'une session, un utilisateur va acheter un produit ou non afin de sélectionner une offre marketing selon le cas.

Phase d'expérimentation
- Prise en main (Récupération des données et Étude descriptive)
- Modélisation
- Google Cloud Platform, Cluster Spark avec Dataproc, Chargement vers BigQuery(50gb)

Software Engineering
- Architecture du pipeline ML (Linting et refactoring PEP 8, Kedro, Versioning avec git)
- Tests logiciels et dépôt de modèles Déploiement de modèles
- API REST (avec FLASK)
- Conteneurisation et serverless (Docker, automatisation de l'exécution de l'API)

MLOps
- Déclencheurs et pipelines CI/CD (Serverless, Conteneurs, …)
- Kubernetes (K8s, ReplicaSets et Deployment, dashboard administrateur)
- Pipeline de production (logging sous Python, Intégrer Cloud Logging…)
- Automatisation (Airflow, DAG, …)
- Sécurité et monitoring (Identification, autorisation et authentification)

Environnement technique : Python, Google Colab, Google cloud platform, Dataproc, bucket, Cloud Storage, Apach spark, spark SQL, PySpark, BigQuery, kedro, MLflow, Serverless, Docker, Kubernetes, Apache Airflow

Expérience professionnelle
KANTAR - 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 puis vers R**
- Prise de connaissance des SPEC Excel et réalisation de nouvelles SPEC, traduites ensuite en programme WPS et SAS
- Migration des programmes de SAS vers WPS.
- Réentrainement 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'entrainement a été réalisé puis présenté avant la mise en production (langage R et Python)

Environnement technique : SAS, WPS, Python, R

EDF LAB R&D - 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)
- Rapprochement par la distance au « sphère de vibration du noyau » (méthode imaginée et modélisée)
- 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 sur base Oracle

IRI-FRANCE Chambourcy, Data Scientist
05/2017 – 08/2017

Projet 1 (data analyst) : Analyse de données ; langages SAS et R.

Projet 2 (data scientist) : Conception et intégration d'un programme d'intelligence artificielle (machine learning) réalisant des estimations.

Mission :

- Prise en main de l'ancien process : langages SAS, Shell-bash, SQL et environnement Unix
- Prédiction de variable quantitative (langage R) : régression linéaire multiple, régression PLS, réseau de neurone, arbre de régression, régression polynomial, régression svm_lineaire
- Création d'un programme de lancement avec demande de profil et mot de passe. Le programme enregistre les tâches, l'heure, le nom du PC et le profil de l'utilisateur dans un fichier " .log". Objectif : faciliter le suivi de l'exécution des tâches par un superviseur d'équipe : langage utilisé R, CMD-Batch et environnement Windows.
- Rédaction de rapports et présentation des résultats

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

Projets personnels & réalisations

2021 – 2022

Data scientist (Machine learning, IA)

Trading automatique: Prédiction et optimisation des prises de position des cours boursiers

Méthode 1 : approche prédictive supervisée

- Détermination des points de position avec un algorithme pour créer la variable cible
- Extraction et sélection des features à l'aide des indicateurs techniques
- Deep learning (tensorflow, keras) pour la prediction
- Optimisation des hyperparamètres
- Validation du modèle et Mise en production sur Systemd de linux (ubuntu)

Méthode 2: IA de reinforcement learning (q-learning)

- Formalisation de l'agent IA et Entrainement de l'agent et mise à jour des résultats dans la table Q
- Établir les états initiaux
- Formalisation des différentes actions possibles,
- Fonction de transmission (Modèles de prédiction en réseau de neurone multicouche) et de récompense
- Choix du meilleur model
- Pipeline de Mise en production de l'agent correspondant au meilleur model sur Système linux (ubuntu)

Développeur Full Stak

Réalisation 1 - Conception de site e-commerce back-end et front-end

- front-end (Django, html, CSS, javascript):
  - Template, intégration API stripe
- Back-end (python, shell Unix):
  - Data base : PostgreSQL
  - Mailing automatique

Réalisation 2 - site web de service numérique

- front-end (Django, html, CSS, javascript):
  - Template, intégration API Stripe
- Back-end (python, shell Unix):
  - Data base : PostgreSQL
  - Web scraping, controle de browser avec selenium,
  - Requests des ulr avec des méthodes POST et Get,
  - création automatique de rapport en pdf et mailing automatique
  - traitement automatique du processus de signature électronique
  - Les jobs sont mis en production dans des containers sur Docker

Others Link [https://github.com/AinaKIKISAGBE/**]:**

- Detection de fraude (secteur bancaire) [lien]
- Detection d'anomalies (industrie) [lien]
- Création de package Python [lien]

Savoir-être

- *Bon relationnel*
- *intéressé par du foot de salle avec mes collègues : 10 au minimum (5 vs 5)*
- *mes collègues préfèrent les AfterWork du Jeudi soir, moi je préfère ceux du Vendredi soir, au final on s'adapte selon nos agendas.*
- *J'aime jouer aussi au jeu de dame*