



Okan KOCABIYIK

Ingénieur IA & Data

Big Data | Machine Learning | IA Générative | MLOps

contact@okandata.com

06.40.49.47.95

LinkedIn : [Okan Kocabiyik](#)

SiteWeb : <https://okandata.com/>

GitHub : <https://github.com/okankcb>

DagsHub : <https://dagshub.com/okankcb>

PROFIL

Spécialisé en modélisation prédictive, industrialisation ML et IA générative. Maîtrise des projets E2E : de l'ingestion des données jusqu'au déploiement cloud (AWS EC2/ECR, CI/CD). Capacité à traiter des volumes massifs (+10 Mds de lignes) avec un fort impact métier mesurable.

COMPÉTENCES CLÉS

- › Modélisation prédictive & scoring
- › Machine Learning (classif., régression)
- › Feature engineering avancé
- › Industrialisation MLOps (CI/CD)
- › IA Générative (RAG, LLM)
- › Big Data >10 Mds lignes
- › Optimisation SQL massive
- › Pipelines ETL/ELT robustes
- › Data Quality & tests auto
- › Architecture E2E : ingestion → déploiement

STACK TECHNIQUE

Langages : Python, SQL, R, Bash

ML/DS : Scikit-learn, XGBoost, LogReg, Random Forest, ACP, KNN, Tensorflow

IA Générative : LangChain, FAISS, Chroma, HuggingFace, OpenAI API, RAG pipelines

Big Data : Vertica, VerticaPy, MongoDB, PostgreSQL

Data Eng. : ETL, MLOps, MLflow, CI/CD, Docker, pipelines batch, DagsHub, Airflow

Cloud : AWS (EC2, ECR, S3, SageMaker)

API/Deploy : FastAPI, Uvicorn, Github Actions

Outils : Git, GitHub/Lab, Jupyter, VS code, conda/mamba, uv

BI : Power BI

SOFT SKILLS

- › Esprit analytique
- › Orientation résultats métier
- › Communication transverse
- › Esprit d'équipe
- › Accompagnement utilisateurs

LANGUES

Français : Langue maternelle

Turc : Langue maternelle

Anglais : Intermédiaire (B1/B2)

Espagnol : Notions (A1)

MOBILITÉ

Rennes / Nantes (2j TT)

Paris (min. 3j TT)

EXPÉRIENCES PROFESSIONNELLES

EURO-INFORMATION – Filiale tech. Crédit Mutuel / CIC | 11/2022 – 03/2026

Ingénieur IA & Data

- Conception d'algorithmes Python pour le scoring, la segmentation et la détection de risques financiers.
- Industrialisation de modèles ML en environnement Big Data (packaging, automatisation, mise en production).
- Recette et montée de version VerticaPy (0.11.x → 1.0.x) : scripts de tests unitaires, contrôles qualité et non-régression.
- Développement d'un modèle prédictif anticipant la fragilité financière à 6 mois (LogReg, XGBoost, Random Forest) — AUC et recall orientés métier.
- Feature engineering avancé : rolling averages, variations dynamiques, signaux faibles, gestion du déséquilibre de classes.
- Analyse de l'impact de l'inflation sur +10 milliards de transactions clients — segmentation comportementale et catégorisation des dépenses.
- Automatisation des calculs d'empreinte carbone (PCAF) pour les reportings ESG réglementaires.
- Contribution aux dispositifs AML : développement et validation de modèles de détection de transactions suspectes.
- Mise en place de pipelines RAG/LLM (LangChain, FAISS, HuggingFace) pour des cas d'usage d'IA générative.
- Support et accompagnement des utilisateurs DataLab (adoption, bonnes pratiques, restitution métier).

BANQUE POPULAIRE GRAND OUEST | 10/2021 – 06/2022

Data Analyst / Data Engineer

- Conception et développement de tableaux de bord Power BI pour le pilotage opérationnel et stratégique.
- Implémentation de requêtes SQL complexes (jointures multi-sources, agrégations, optimisation performance).
- Mise en place de stratégies de ciblage client : encodage, génération de fichiers batch et suivi des campagnes.
- Détection et suivi de profils à risque (fraude, incidents, inactivité) via indicateurs analytiques.
- Automatisation des reportings et traitements via — réduction des tâches manuelles.
- Collaboration avec les équipes métiers pour formalisation des besoins et validation des KPIs.

CNP ASSURANCES | 09/2020 – 08/2021

Data Scientist

- Gestion complète d'un projet data science dédiée à l'analyse comportementale des assurés.
- Développement de modèles de scoring (rétention, inactivité) en Python, R et SQL.
- Modélisation actuarielle des impacts du réchauffement climatique sur les lois de longévité.
- Contribution à l'optimisation des campagnes marketing via segmentation et analyses prédictives.

PROJET PERSONNEL

Assistant IA (RAG) : Développement d'un pipeline RAG (LangChain, FAISS, LLM) avec ingestion, vectorisation et déploiement via API (FastAPI).

Pipeline Data & MLOps : Conception d'un pipeline ML industrialisé (ETL, entraînement, déploiement) avec CI/CD, Docker et automatisation des environnements.

FORMATION

Master Ingénierie Économique et Financière

Université de Rennes 1 | 2019 – 2021

Licence Économie-Gestion

Université de Rennes 1 | 2016 – 2019