



DOSSIER DE COMPÉTENCES

Ingénieur IA & Data

Okan KOCABIYIK

Big Data | Machine Learning | IA Générative | MLOps

✉ contact@okandata.com

☎ 06.40.49.47.95

🌐 <https://okandata.com/>

🔗 LinkedIn : [Okan Kocabiyik](#)

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

💻 GitHub : github.com/okankcb

Disponibilité : Immédiate | Télétravail : 2j/sem (Rennes/Nantes) • 3j/sem (Paris)

PROFIL SYNTHÉTIQUE

Fort de plus de 5 ans d'expérience dans l'écosystème data et intelligence artificielle, je suis un ingénieur IA & Data polyvalent, spécialisé dans la conception et l'industrialisation de solutions à fort impact métier. Mon parcours dans des environnements exigeants et à grande échelle m'a permis de développer une expertise combinant rigueur analytique, maîtrise des volumes massifs de données et capacité à délivrer des modèles en production.

Je maîtrise le cycle complet d'un projet data science : de l'ingestion des données brutes jusqu'au déploiement cloud via des pipelines CI/CD robustes, en intervenant aussi bien sur les phases amont (qualification des données, KPIs métier) que sur les phases aval (packaging, automatisation, monitoring).

J'ai développé et industrialisé des modèles de scoring et de classification complexes, notamment des modèles de prédiction comportementale à 6 mois (LogReg, XGBoost, Random Forest), optimisés sur l'AUC et le recall métier. Mes compétences en feature engineering avancé me permettent d'extraire la valeur maximale de données hétérogènes, y compris sur des volumes dépassant 10 milliards de lignes.

J'ai mis en place des pipelines RAG (LangChain, FAISS, HuggingFace) intégrant l'IA générative dans des architectures de production, et j'applique systématiquement les bonnes pratiques MLOps : MLflow, Docker, GitHub Actions et AWS pour des déploiements scalables et reproductibles.

Mes interventions couvrent des domaines variés : scoring de risque, détection d'anomalies, reporting réglementaire, analyse comportementale et segmentation client. Reconnu pour ma capacité à communiquer avec les équipes métier, je sais formaliser les besoins, restituer les résultats et accompagner l'adoption des outils data.

COMPÉTENCES TECHNIQUES

Langages	Python · SQL · R · Bash
ML / Data Science	Scikit-learn · XGBoost · LogReg · Random Forest · KNN · ACP · TensorFlow
IA Générative	LangChain · FAISS · Chroma · HuggingFace · OpenAI API · RAG pipelines
Big Data	Vertica · VerticaPy · MongoDB · PostgreSQL
Data Engineering	ETL/ELT · MLOps · MLflow · CI/CD · Docker · Pipelines batch · DagsHub
Cloud	AWS (EC2, ECR, S3, SageMaker)
API / Déploiement	FastAPI · Uvicorn · GitHub Actions
Outils	Git · GitHub/Lab · Jupyter · VS Code · conda/mamba · uv
BI	Power BI

NIVEAU D'EXPERTISE PAR DOMAINE

Domaine	Niveau	Évaluation
Modélisation ML / Scoring	Senior	★★★★★
Feature Engineering	Senior	★★★★★
MLOps / Industrialisation	Confirmé	★★★★☆
IA Générative / RAG / LLM	Confirmé	★★★★☆
Big Data (>10 Mds lignes)	Senior	★★★★★
Cloud AWS	Confirmé	★★★★☆
SQL / Bases de données	Senior	★★★★★
Data Visualisation / BI	Confirmé	★★★★☆

COMPÉTENCES CLÉS MÉTIER

Modélisation & Scoring

- Modèles prédictifs comportementaux
- Scoring et segmentation client
- Détection d'anomalies et de risques
- Classification et régression supervisée

Analyse de données à grande échelle

- Analyse comportementale sur volumes massifs
- Segmentation et ciblage client
- Analyse de l'impact environnemental (ESG / PCAF)
- Traitement de +10 Mds de lignes

IA Générative & NLP

- Pipelines RAG sur bases documentaires
- Recherche sémantique et vectorisation
- Interfaces en langage naturel

Data Engineering & Reporting

- Pipelines ETL/ELT automatisés
- Tableaux de bord et restitution métier
- Automatisation des reportings réglementaires

MLOps & Industrialisation

- Versionnage des expériences (MLflow, DagsHub)
- Conteneurisation et déploiement (Docker, AWS)
- Automatisation CI/CD (GitHub Actions)
- Pipelines d'entraînement et de monitoring
- Packaging et mise en production de modèles

Data Quality & Fiabilité

- Tests unitaires et non-régression
- Contrôle qualité des pipelines de données
- Validation et documentation des modèles
- Gestion de la dette technique data
- Automatisation des contrôles qualité

EXPÉRIENCES PROFESSIONNELLES

EURO-INFORMATION – Filiale tech. Crédit Mutuel / CIC

Ingénieur IA & Data

Nov. 2022 – Mars 2026

Principales réalisations :

- 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)

Technologies : Python · SQL · VerticaPy · Scikit-learn · XGBoost · LangChain · FAISS · HuggingFace · MLflow · Docker · AWS · CI/CD · FastAPI

BANQUE POPULAIRE GRAND OUEST

Data Analyst / Data Engineer

Oct. 2021 – Juin 2022

- 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 — réduction significative des tâches manuelles

- Collaboration avec les équipes métiers pour la formalisation des besoins et la validation des KPIs

Technologies : SQL · Power BI · Python · ETL · Reporting automatisé

CNP ASSURANCES

Data Scientist

Sept. 2020 – Août 2021

- 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

Technologies : Python · R · SQL · Scikit-learn · Modélisation actuarielle · Segmentation

PROJETS NOTABLES

Modèle de Fragilité Financière à 6 Mois

Contexte : Crédit Mutuel / CIC – anticipation des situations de fragilité client

Réalisation : Feature engineering avancé (rolling averages, signaux faibles), comparaison LogReg / XGBoost / Random Forest, gestion du déséquilibre de classes, optimisation métier (AUC, recall).

Impact : Détection précoce des clients fragiles, réduction des risques de défaut, outil d'aide à la décision pour les conseillers.

Stack : [Python](#) · [XGBoost](#) · [Scikit-learn](#) · [Vertica](#) · [MLflow](#) · [Docker](#)

Analyse Inflationniste sur +10 Milliards de Transactions

Contexte : Cartographie de l'impact de l'inflation sur les comportements de dépenses de millions de clients bancaires.

Réalisation : Traitement de volumes massifs en environnement Big Data Vertica, segmentation comportementale multi-critères, catégorisation fine des dépenses par famille.

Impact : Identification de tendances de consommation clés, rapports stratégiques pour la Direction, aide au pilotage commercial.

Stack : [Python](#) · [VerticaPy](#) · [SQL](#) · [Vertica](#) · [Segmentation](#)

Pipeline RAG / IA Générative pour DataLab

Contexte : Mise en place de cas d'usage d'IA générative.

Réalisation : Architecture et déploiement de pipelines RAG (Retrieval-Augmented Generation) avec LangChain, indexation vectorielle FAISS/Chroma, intégration de modèles HuggingFace et OpenAI API.

Impact : Automatisation de la recherche documentaire interne, gain de temps significatif pour les analystes, validation des premiers cas d'usage LLM en production.

Stack : [LangChain](#) · [FAISS](#) · [Chroma](#) · [HuggingFace](#) · [OpenAI API](#) · [Python](#)

Automatisation Empreinte Carbone (PCAF / ESG)

Contexte : Conformité réglementaire ESG et reporting PCAF pour le groupe bancaire.

Réalisation : Développement de scripts Python d'automatisation du calcul de l'empreinte carbone des portefeuilles, mise en place de pipelines batch pour les reportings périodiques.

Impact : Réduction des délais de reporting ESG, fiabilisation des données transmises aux régulateurs.

Stack : [Python](#) · [SQL](#) · [Vertica](#) · [ETL](#) · [Pipelines batch](#)

FORMATION

Master Ingénierie Économique et Financière

Université de Rennes 1 | 2019 – 2021

Licence Économie-Gestion

Université de Rennes 1 | 2016 – 2019

LANGUES

Langue	Niveau	Appréciation
Français	Langue maternelle	★★★★★ Courant
Turc	Langue maternelle	★★★★★ Courant
Anglais	Intermédiaire B1/B2	★★★★☆ Professionnel
Espagnol	Notions A1	☆☆☆☆☆ Débutant

SOFT SKILLS & ATOUTS

Analytique Esprit analytique fort, orientation résultats métier mesurables	Collaboratif Communication transverse, esprit d'équipe, accompagnement utilisateurs	Autonome Gestion E2E de projets complexes en environnement Big Data
---	--	--

MOBILITÉ & CONDITIONS

Zone principale Rennes / Nantes <i>Télétravail : 2 jours/semaine</i>	Zone étendue Paris <i>Télétravail : minimum 3 jours/semaine</i>
---	--