



Okan KOCABIYIK

Data Scientist/Engineer +5 ans XP

BIG DATA | MACHINE LEARNING | IA GENERATIVE

www.okandata.com
[in okan-kocabiyik-630901189](https://in.linkedin.com/in/okan-kocabiyik-630901189)
contact@okandata.com
06.40.49.47.95
Rennes / Nantes (2j TT)
Paris (min. 3j TT)

PROFIL

Data Scientist spécialisé en modélisation prédictive et exploitation de données à très grande échelle. Conception de modèles de scoring, détection de risques et feature engineering avancé, avec une forte orientation impact métier. Industrialisation de pipelines analytiques robustes et scalables, intégrés en production, incluant des projets ESG et IA avancée (RAG, LLM)

COMPÉTENCES CLÉS

- Modélisation prédictive, scoring & détection de risques
- Industrialisation de pipelines analytiques
- Feature engineering
- Big Data (SQL optimisé, >10Md lignes)
- IA générative (RAG, LLM)
- Analyse ESG (PCAF)

STACK TECHNIQUE

- **Langages et Outils** : Python, SQL, R, VerticaPy, Git, GitHub, GitLab, Jupyter, VS Code, conda/mamba, Office 365
- **Data & Big Data** : Vertica optimisation SQL (volumes massifs)
- **Machine Learning** : Scikit-learn, XGBoost, régression, classification, évaluation modèle
- **IA Générative** : LangChain, FAISS, Chroma, HuggingFace, OpenAI API, RAG pipelines
- **Data Engineering** : ETL, pipelines batch, industrialisation, CI/CD

SOFTS SKILLS

- Esprit analytique
- Esprit d'équipe
- Communication transverse
- Accompagnement métier

Langues

- Français : Langue maternelle
- Turc : Langue maternelle
- Anglais : Niveau intermédiaire
- Espagnol : Niveau débutant

FORMATIONS

- ❖ Master Ingénierie Economique et financière | 2019-2021 | Université de Rennes 1
- ❖ Licence Economie-Gestion | 2016-2019 | Université de Rennes 1

Expériences Professionnelles

- 🚩 **EURO – INFORMATION (Filiale technologique Crédit Mutuel – CIC) | 11/2022 – 03/2026 |**
 - Conception d'algorithmes Python pour le scoring, la segmentation et la détection de risques.
 - Industrialisation des modèles en environnement Big Data (automatisation, packaging, mise en production).
 - Recette VerticaPy (0.11.x → 1.0.x) avec test unitaires & Développement de scripts et bibliothèques Python pour automatiser les recettes VerticaPy (tests data, contrôles qualité, non-régression) et fiabiliser les déploiements.
 - Développement d'un modèle prédictif anticipant la fragilité financière à 6 mois.
 - Feature engineering avancé et gestion du déséquilibre de classes (LogReg, XGBoost, Random Forest).
 - Analyse de l'impact de l'inflation sur +10 milliards de transactions clients.
 - Segmentation comportementale et catégorisation avancée des dépenses.
 - Conception d'une table de catégorisation des dépenses (logement, alimentation, santé, loisirs...) via règles métiers et nouvelles nomenclatures.
 - Feature engineering avancé (rolling averages, variations dynamiques, événements récents) pour capter les signaux faibles.
 - Validation via métriques orientées métier (AUC, recall, taux de détection anticipée) et comparaison aux modèles existants.
 - Automatisation des calculs d'empreinte carbone dans le cadre des reportings ESG (PCAF).
 - Contribution aux dispositifs AML avec développement et validation de modèles de détection de transactions suspectes.
 - Support & accompagnement des utilisateurs DataLab (adoption, bonnes pratiques, restitution).
- 🚩 **BANQUE POPULAIRE GRAND OUEST | 10/2021 – 06/2022 |**
 - Conception et développement de tableaux de bord Power BI pour le pilotage opérationnel et stratégique.
 - Implémentation de requêtes SQL complexes pour l'extraction, la préparation et l'analyse de données multi-sources.
 - Mise en place de stratégies de ciblage client : encodage des données, 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 des traitements via Visual Tom, réduisant les tâches manuelles et améliorant la fiabilité des données.
 - Collaboration avec les équipes métiers pour formalisation des besoins et validation des KPIs.
- 🚩 **CNP ASSURANCES | 09/2020 – 08/2021 |**
 - Gestion complète d'un projet de data science dédié à l'analyse comportementale des assurés.
 - Développement de modèles de scoring (rétention, inactivité) en R et SQL.
 - Évaluation et comparaison des performances via métriques statistiques adaptées.
 - Modélisation actuarielle des impacts du réchauffement climatique sur les lois de longévité.
 - Contribution à l'optimisation des campagnes marketing via segmentation et analyses prédictives.