



OKAN KOCABIYIK

Data Scientist



Rennes



contact@okandata.com



[GitHub](#)



06.40.49.47.95



[LinkedIn](#)



[SiteWeb](#)

Compétences Techniques & Transverses

- **Data Science & Machine Learning** : modélisation prédictive, scoring client, détection d'anomalies, analyse comportementale
- **Big Data** : SQL optimisé, Vertica/VerticaPy, traitement de 10+ milliards de lignes
- **Développement** : Python, R, SQL, pipelines, scripts, industrialisation
- **Mise en production** : packaging, automatisation, intégration métier
- **Optimisez avec l'automatisation** : Automatiser des tâches répétitives grâce à des workflows simples et efficaces adaptés à l'activité
- **Déployez une architecture data** robuste avec pipelines, stockage optimisé et processus de nettoyage automatisés
- **Data Analysis & BI** : KPIs, dashboards Power BI, reporting automatisé
- **Qualité & Gouvernance des données** : documentation, standardisation
- **Recueil des besoins métier** : cadrage, restitution, accompagnement décisionnel
- **Esprit analytique** : orientation impact métier, transformation des données en actions concrètes
- **Communication transverse** : collaboration avec Data Architects, Métiers, DevOps, Risk Officers, DS
- **Outils IA & automatisation (formation et veille)** :
 - **RAG (Retrieval-Augmented Generation)** : compréhension du fonctionnement des frameworks LangChain, LangGraph, LangSmith, création de bases vectorielles (FAISS, Chroma), et structuration de documents pour exploitation IA.
 - **HuggingFace Transformers / SentenceTransformers** – embeddings, NLP avancé, modèles de texte.
 - **PyMuPDF, pdfplumber, BeautifulSoup, docx2txt** – parsing documentaire et extraction de contenu pour IA.
 - **Tesseract, EasyOCR, PaddleOCR** – reconnaissance optique de caractères (OCR).
 - **LayoutLM, Donut, BLIP, GPT-4V** – modèles Vision-Langage (VLM) appliqués à la compréhension de documents visuels.
 - **OpenAI API / Mistral / Claude** – expérimentation sur la génération de contenu et l'automatisation intelligente via LLMs.

Logiciels & Langages

- **Python** – data science, data engineering, data analyst, automatisation de traitements
- **R** – modélisation statistique, actuariat, analyses comportementales
- **SQL** – extraction, transformation, manipulation de données sur bases massives

- **DAX** – calculs personnalisés dans Power BI (KPIs, ratios dynamiques, filtres de contexte)
- **Power Query** – transformation de données dans Power BI
- **VBA** – automatisation dans Excel, génération de rapports
- **Shell (Bash, Zsh)** – scripting, automatisation sous Unix/Linux
- **C++** – notions de base, utilisé en contexte académique ou projets spécifiques
- **Power BI Desktop & Service** – tableaux de bord interactifs et dynamiques
- **Microsoft Office 365** – Excel, Word, PowerPoint, Outlook, One Note, One Drive
- **Jupyter Notebook / JupyterLab** – développement Python / R interactif
- **VS Code** – environnement de développement multi-langages
- **Git / GitHub / GitLab** – gestion de version, CI/CD
- **VTOM** – supervision et ordonnancement de traitements batch
- **Conda / Mamba** – gestion d'environnements Python/R
- **Notepad++ / Sublime Text**

Formation & Langues

Formation	BAC \geq 5 – Master Ingénierie Economique et Financière parcours Data Science (Université de Rennes 1)	2021
Langues	<ul style="list-style-type: none"> • Anglais - Intermédiaire • Français – Natif • Turc – Natif • Espagnol – Débutant 	

EXPERIENCES PROFESSIONNELLES

Euro – Information

30 MOIS (11/2022)

Data Scientist / Data Engineer

Contexte

Mon rôle a consisté à accompagner les métiers dans l'exploitation de la donnée, à concevoir des solutions analytiques et prédictives à grande échelle, et à industrialiser les traitements dans un environnement Big Data, en collaboration avec les Data Architects, Data Scientists et Data Officers.

Activités : Pilotage et Développement en Data Science

- Accompagnement des métiers dans l'intégration de la data science dans leurs processus décisionnels : cadrage des besoins, définition des objectifs analytiques.
- Conception, développement et optimisation d'algorithmes en Python et R pour la modélisation prédictive, le scoring client, la segmentation et la détection de comportements à risque.
- Industrialisation des traitements analytiques dans un environnement Big Data : packaging de modèles, automatisation des flux, intégration dans les chaînes de production.
- Déploiement de VerticalPy (recette des versions 0.11.x à 1.0.x) et développement de librairies personnalisées pour l'automatisation des recettes de données.
- Contribution à la stratégie d'innovation data en lien avec les Data Architects et Data Officers, dans une logique d'industrialisation et de gouvernance des modèles.
- Accompagnement des utilisateurs DataLab

Activités : Élaboration d'un modèle prédictif – Prédiction des tiers fragilisés financièrement à 6 mois

- Objectif du projet : anticiper les situations de fragilité financière des clients du groupe bancaire à un horizon de 6 mois, afin de déclencher des actions préventives ciblées.
- Construction d'un jeu de données temporel, intégrant des historiques de comportements financiers (dépenses, encours, incidents de paiement, revenus, aides sociales, etc.) sur plusieurs années.
- Développement de variables explicatives temporelles (rolling averages, variations mensuelles, événements récents) pour capter les signaux faibles d'une détérioration de la situation financière.
- Modélisation supervisée via des algorithmes de type régression logistique, XGBoost et Random Forest, en intégrant des techniques de gestion du déséquilibre de classes.
- Validation du modèle avec des métriques orientées métier (précision, rappel, AUC, taux de détection anticipée) et comparaison avec des modèles baselines existants.
- Mise en place d'une logique de scoring périodique, destinée à être intégrée dans un outil de pilotage pour les services de prévention et d'accompagnement client.
- Restitution des insights clés auprès des directions Conformité et Risques, avec un focus sur les principales variables contributives à la prédiction de la fragilité.

- Rédaction d'un cahier des charges fonctionnel et d'une documentation statistique détaillée, incluant la description des variables, la méthodologie de modélisation, les choix techniques et les résultats des évaluations.

Activités : Analyse de l'impact de l'inflation sur les clients particuliers

- Objectif du projet : étudier l'effet de l'inflation sur les comportements de consommation des retraités et identifier les populations les plus exposées à la perte de pouvoir d'achat.
- Collaboration avec le Conseil d'Analyse Économique (CAE) pour définir les axes d'analyse, les catégories de dépenses sensibles et les indicateurs économiques à suivre.
- Traitement d'un historique de données bancaires sur 3 ans, incluant plus de 10 milliards d'enregistrements de transactions clients.
- Création d'une table de catégorisation des dépenses, basée sur des règles internes et de nouvelles nomenclatures, permettant de regrouper les opérations par grandes familles de consommation (logement, alimentation, santé, loisirs...).
- Analyse d'évolution de comportements : détection de changements significatifs dans les habitudes de dépenses, segmentation des retraités selon leur profil d'adaptation à l'inflation.

Activités : Projet de Lutte contre le Blanchiment d'Argent

- Objectif du projet : renforcer les capacités de détection et d'analyse des transactions suspectes en lien avec des opérations de blanchiment d'argent, en appui aux équipes Conformité et Inspection Générale.
- Collaboration étroite avec l'Inspection Générale : recueil et formalisation des besoins métier via l'élaboration d'un cahier des charges précis, incluant contraintes réglementaires et exigences de traçabilité.
- Itérations successives entre modélisation, validation métier et calibration du modèle, avec restitution des résultats sous forme de rapports interactifs et exportables.

Activités : Automatisation des recettes VerticaPy (librairie Python)

- Mise en place d'un algorithme permettant d'automatiser l'exécution des notebooks VerticaPy dans le cadre des recettes techniques.
- Cette automatisation permet un gain de temps significatif et une standardisation des contrôles.
- Le lancement séquentiel ou parallèle de plusieurs notebooks,
- La génération automatique d'un fichier de synthèse indiquant le statut d'exécution de chaque notebook
- L'intégration de tests unitaires avec reporting des résultats pour chaque étape critique.

Activités : Projet ESG, Automatisation des calculs de l'empreinte carbone des financements

- Contribution directe aux reportings réglementaires ESG et au suivi PCAF du Crédit Mutuel.
- Mise en place de pipelines robustes et automatisés permettant de fiabiliser et d'accélérer les

calculs d'empreinte carbone.

- Diagnostics sur la qualité et la stabilité des codes.

Environnement Technique : Office 365, Jupyter LAB, Git, GitHub, GitLab, SQL, Verticapy, Python, R, Unix/Linux, Conda / Mamba, Utilisation avancée du terminal (Bash, Zsh)

Contexte

Développement de solutions analytiques et de tableaux de bord pour le suivi de l'activité, la détection des profils à risque, et l'optimisation des reportings métiers. Le projet visait à automatiser les traitements de données, à améliorer la qualité et la fiabilité des indicateurs, et à fournir aux équipes métiers des outils d'aide à la décision en temps réel, notamment via Power BI.

Activités : Développement et Implémentation de Solutions Data

- Conception et exécution de requêtes SQL complexes sous SAS pour l'extraction, la préparation et l'analyse de données multi-sources.
- Mise en place de stratégies de ciblage client, incluant l'encodage des données et la génération de fichiers batch pour le suivi des réseaux et des campagnes marketing
- Développement de tableaux de bord dynamiques sous Power BI, adaptés aux besoins métiers (commerciaux, conformité, risques).
- Refonte de reportings existants pour les rendre plus clairs, interactifs et automatisés, en réduisant les erreurs manuelles.
- Participation active à la formalisation des besoins métiers, à la validation des indicateurs, et à la documentation des traitements analytiques.

Activités : Surveillance et Gestion des Risques

- Mise en place de processus d'identification et de suivi des clients à risque (fraude, incidents de paiement, inaktivité), à partir de données transactionnelles et comportementales.
- Calcul de KPIs, statistiques descriptives
- Création de dashboards Power BI dédiés au pilotage des risques, permettant une visualisation claire des alertes, tendances et zones critiques.
- Appui aux équipes métiers dans l'interprétation des données de risque et la mise en œuvre d'actions correctives.

Activités : Crédit et Automatisation de reportings

- Conception et développement de reportings automatisés à destination des équipes commerciales, des directions métiers et des clients externes. Crédit et automatisation de la production des tableaux de bord pour réduire les tâches manuelles et accélérer la diffusion des résultats.
- Refonte de reportings existants afin d'améliorer la lisibilité, la fréquence de mise à jour et la pertinence des indicateurs suivis
- Développement de nouvelles solutions de reporting sous Power BI, intégrant des filtres dynamiques, des agrégations multi-niveaux et des alertes visuelles.
- Automatisation de la production des tableaux de bord pour réduire les tâches manuelles et accélérer la diffusion des résultats.
- Mise en place de standards de reporting pour assurer l'uniformité des livrables et faciliter l'appropriation des outils par les utilisateurs métiers.

CNP Assurances

12 MOIS

(09/2020)

Data Scientist

Contexte

Concevoir, développer et déployer des modèles de scoring client et de modélisation actuarielle, avec un focus sur l'analyse de l'impact du réchauffement climatique sur les lois de longévité et la comportementalisation des assurés. Ces modèles devaient être intégrés aux campagnes marketing multicanal et aux outils de pilotage client.

Activités : Conception et automatisation de modèles

- Développement de modèles de scoring de rétention et d'inactivité à l'aide de R et SQL.
- Mise en Place et Analyse des Modèles : Déploiement des modèles de scoring dans diverses campagnes marketing, avec une analyse rigoureuse de leur performance.

Activités : Analyse comportementale & statistique

- Réalisation d'analyses descriptives pour identifier des profils types et comportements d'adhésion/résiliation.
- Modélisation statistique pour la segmentation client et l'optimisation des campagnes.

Environnement Technique : Office 365, Python, R, SQL