

SINHOU MAGNUS

Data Scientist - Alternance

0615409515 · reubenmag6@gmail.com · sinhoumagnus.fr · github.com/reubenmag

Étudiant en fin de Bachelor Data & IA, je conçois des pipelines ML end-to-end - du prétraitement à l'évaluation - sur des données réelles (images, séries temporelles, transactions). Solide en Python, scikit-learn et statistiques inférentielles. Je recherche une alternance Data Scientist pour septembre 2026.

EXPÉRIENCE

Data Analyst - Stage · ARKOSE, Montreuil (Paris) · 2 mois · 2025

- Pipeline de nettoyage et validation de 158 000+ enregistrements (Python, pandas) avec recalcul des indicateurs clés
- Feature engineering : cumuls mensuels, ratios, détection de saisonnalité et profils comportementaux (fidèles / occasionnels / nouveaux)
- Production de recommandations marketing data-driven et tableau de bord de suivi KPIs

PROJETS

Détection de canalisations souterraines - ML sur données drone - données réelles · partenariat entreprise · 2026

- Pipeline de chargement et prétraitement de scans multi-canaux (.npz) acquis par drone sur infrastructure réelle : resize 224×224, gestion des NaN
- Réduction dimensionnelle par PCA 2D pour visualisation et préparation des features
- Optimisation du KNN par GridSearchCV + StratifiedShuffleSplit (10 folds) - F1-score > 99% en validation croisée
- Diagnostic du domain gap entre données synthétiques et réelles ; analyse des limites de généralisation
- Projet mené sur données terrain fournies par une entreprise partenaire - confrontation directe aux problématiques de qualité et de variabilité des données réelles

Classification d'images médicales - Détection du paludisme - académique · 2025

- Pipeline de classification d'images de cellules sanguines (saines / infectées) sur Google Colab
- Prétraitement, augmentation de données et entraînement d'un modèle CNN (PyTorch/Keras)
- Structure modulaire et réutilisable pour faciliter l'expérimentation

Analyse forensique financière - Affaire Mercier & Associés - personnel · avril 2026

- Exploration de 5 jeux de données : 2 500 transactions, 10 000 logs d'accès, 300 mesures biométriques - détection de virements offshore frauduleux
- Tests statistiques inférentiels (Shapiro-Wilk, t-test, ANOVA, Fisher) + ML non supervisé (PCA, K-Means) et supervisé (KNN, arbre de décision)
- Classification automatique de 1 000 emails suspects via API Anthropic (LLM) - restitution en rapport forensique structuré

COMPÉTENCES TECHNIQUES

Langages : Python · SQL · DAX

ML / Stats : scikit-learn · PCA · KNN · Random Forest · K-Means · Tests inférentiels · Feature Engineering

Data & BI : pandas · NumPy · Power BI · Google BigQuery · PostgreSQL · MongoDB

Cloud & Outils : Amazon AWS · GitHub · API Anthropic (LLM)

FORMATION

Bachelor Data & IA · HETIC · 2024 - 2026

Baccalauréat Scientifique · 2023

LANGUES & CENTRES D'INTÉRÊT

Langues : Français (natif) · Anglais (professionnel)

Intérêts : Basketball · Football · Littérature financière