Thibault CLEMENT

Data Scientist/Engineer
Freelance

thibault.clement@intechnia.fr
www.intechnia.fr
t-clement
thclement

Compétences clés

Dev : Python, R, Spark, git, bash
Data : SQL, No-SQL, S3, Postgre, ETL
Viz : PowerBi, Qlikview, QGis
Ops : Airflow, CI/CD
Cloud : Amazon Web Services (AWS)
Web : Flask, Streamlit, FastAPI

Expériences Professionnelles

Depuis 2025 **Data Engineer**, *Engie Green*, Montpellier

Récupération et analyse des données historiques des actifs renouvelables

- Conception et développement d'un **pipeline de parsing et d'intégration de données hétérogènes internes** (Python, Airflow, AWS, Flask), garantissant la collecte et l'unification des données
- **Analyse avancée des courbes de charge** et contrôle de la qualité des données temporelles pour détecter et corriger les incohérences
- Création d'un programme de validation automatisée pour **identifier les anomalies** et les données manquantes, assurant une qualité de données optimale
- Développement et mise en place de **tableaux de bord de contrôle** (Power BI, Flask), permettant un suivi en temps réel de la qualité des données
- Optimisation des performances du pipeline pour garantir un **traitement fiable, scalable et automatisé** des données à grande échelle

2021 – 2023 **Data Scientist - Data Engineer**, *Edater*, Montpellier

Modélisation de l'impact économique du secteur maritime

- Classification des descriptions d'activité via **NLP**, en collaboration avec le client, améliorant la précision de la classification NAF
- Estimation du chiffre d'affaires et de la valeur ajoutée des entreprises à partir des données **URSSAF** et **MSA**
- Déploiement multi-régional : **Nouvelle-Aquitaine** (2021), **Occitanie** (2022)

Analyse surfacique des territoires agricoles en Bretagne (2010-2020)

- Accréditation pour le traitement de données sensibles sur un environnement sécurisé **CASD**, analyse du recensement agricole individuel
- Analyse contrefactuelle des exploitants agricoles selon leur accès aux régimes d'aides

Analyse temporelle des trajectoires des entreprises défaillantes

- **Rétro-ingénierie** des schémas de défaillance pour identifier les motifs récurrents précédant la défaillance
- Modélisation prédictive des entreprises à risque en intégrant des variables temporelles, sectorielles et géographiques

Création d'un système d'information (SI) sur 5M d'entreprises

- Développement d'un **pipeline de traitement mensuel** de données (activité, emploi, GPS, finance) à partir de sources open data (**SIRENE, BODACC, RNCS, RNM, RNA**)
- Conception d'un algorithme d'appariement optimisé pour retrouver les entreprises par raison sociale (**+300% de rapidité** grâce à l'optimisation et à la parallélisation)

2018 – 2020 **Data Analyst - Responsable Applicatif**, *CDC Habitat*, Montpellier, TMA via Business & Decision

- Migration du reporting de Hummingbird BI vers QlikView, avec conception et déploiement de tableaux de bord stratégiques (patrimoine, achats, ventes, CRM, etc.)
- Développement et maintenance évolutive pour la centralisation et l'optimisation des services SI dans le cadre du projet **Trajectoire 2022**
- Gestion des demandes et incidents via **Jira**, documentation fonctionnelle et technique sur **Confluence**, et suivi du versioning avec **Git/Bitbucket**
- Automatisation des rafraîchissements mensuels des données et contrôle qualité des flux
- Supervision et résolution des incidents liés aux reportings sous **SAP BO**

2018 – 2018 **Économètre**, *Paris School of Economics (PSE)*, Paris

Pilote d'une expérience en choix discrets dans le cadre du projet AMBU-ENDO (Hospinnomics)

- Conception et analyse d'un modèle de choix discret (DCE) pour évaluer les préférences des professionnels de santé entre les modes de prise en charge ambulatoire et traditionnel du cancer de l'endomètre
- Utilisation de modèles probit et logit mixte pour analyser les critères de faisabilité, sécurité, qualité de vie et coût
- Coordination de la collecte de données et des analyses statistiques pour fournir des recommandations pour les politiques de santé

2017 – 2017 **Économètre**, *CHU Montpellier - DIM*, Montpellier

Étude économétrique sur l'impact du lymphœdème sur la qualité de vie

- Traitement et analyse d'une base de données longitudinale sur 2 ans, incluant des mesures de bien-être (EQ-5D-3L, ICECAP-A, LYMQOL)
- Étude de la complémentarité et de la concordance des indicateurs via matrices de corrélation, diagrammes de dispersion et graphiques de Bland-Altman
- Application d'analyses factorielles (ACP, analyse discriminante) pour identifier les relations entre les dimensions du bien-être
- Exploration de modèles économétriques (logistique ordinale) pour modéliser les comportements des patients face aux choix de réponse

Formations

2024 – 2024 **Certificate of Achievement C1**, *EF School*, Manchester (R-U), C1 level
Programme d'immersion intensive, avec des cours théoriques et pratiques pendant deux mois

2020 – 2021 **D.U. - Data Science & Big Data**, Université de Montpellier, mention TB
Python, TensorFlow, SQL, R, machine learning, deep learning, actuariat

2017 – 2025 **MOOC**, FUN, Coursera, Udemy, DataCamp
Formation en continue : Python, SQL, R, Spark, AWS, NLP, Docker, etc.

2016 – 2018 **Master d'Économie**, *Spécialité comportementale et décisionnelle*, Université de Montpellier, mention B
Économétrie, mathématiques, théorie des jeux, macroéconomie, microéconomie, optimisation dynamique

2012 – 2016 **Licence d'Économie**, *Économie-Gestion*, Université de Bourgogne, Dijon
Économétrie, mathématiques, statistiques, macroéconomie, microéconomie

2008 – 2012 **Baccalauréat ES**, *Économique et Social*, Lycée Stephen Liégeard, Brochon, mention AB
Spécialité mathématiques

Projets personnels

Depuis 2024 **Observatoire des emplois dans la data**, *Mise en place d'une veille automatisée du marché de la data, avec analyse de l'offre et de la demande pour un positionnement stratégique optimal*
Python, web scraping, scikit-learn, ACP

Depuis 2022 **Prédiction des résultats de courses hippiques**, *Développement de modèles avancés de Machine Learning et Deep Learning pour la prédiction des résultats des courses hippiques, avec une gestion automatisée de bankroll optimisée visant une rentabilité positive et durable*
Python, scikit-learn, gestion de bankroll, statistiques, économétrie

Depuis 2020 **Site web personnel**, *Conception, hébergement et administration d'un blog et portfolio personnel sur un serveur auto-hébergé, servant de vitrine pour mon activité de Data Scientist freelance*
Linux, Raspberry Pi, Hugo, Markdown, ssh, darkhttpd, stunnel

Concours

2020 **BFore.ai**, *ISDM Montpellier*, *#1 Première place*
Développement d'un modèle de détection d'adresses IP frauduleuses via un stacking de Random Forest optimisés, sur des données réelles anonymisées fournies par un organisme de cybersécurité

2020 **Montpellier Méditerranée Métropole**, *ISDM Montpellier*, *#2 Seconde place*
Développement d'un modèle de stacking XGBoost optimisé pour estimer les prix des biens immobiliers selon différents critères, en exploitant des données locales et des variables socio-économiques