



**THOMAS MEBARKI**



INGÉNIEUR MACHINE LEARNING

# Dossier de compétences

**Thomas Mebarki**

Ingénieur en Machine  
Learning - Lille, France

- **Tel :** [+33622515934](tel:+33622515934)
- **Email :**  
[thomas.mebarki@protonmail.com](mailto:thomas.mebarki@protonmail.com)
- **LinkedIn :**  
<https://www.linkedin.com/in/thomasmeb>
- **GitHub :**  
<https://github.com/ThomasMeb>

---

## Résumé Professionnel

Ingénieur en machine learning avec une expérience variée dans des startups et des projets personnels couvrant la modélisation prédictive, la détection d'anomalies, et la classification d'images. Spécialisé dans la création de modèles performants et interprétables, notamment pour le secteur bancaire et la détection de fraudes. Passionné par les nouvelles technologies et l'intelligence artificielle, je suis toujours à la recherche de défis pour transformer des données en valeur ajoutée.

## Compétences Techniques

- **Langages** : Python, SQL
- **Machine Learning** : Classification, régression, détection d'anomalies, modélisation prédictive
- **Algorithmes** : Random Forest, XGBoost, LSTM, GARCH, Isolation Forest, CNN (VGG16, ResNet, EfficientNet)
- **Outils et Bibliothèques** : Scikit-Learn, TensorFlow, Keras, Apache Kafka, RabbitMQ
- **Prétraitement des données** : Lemmatisation, tokenisation, PCA, data augmentation
- **Métriques d'évaluation** : AUC-ROC, F1-score, RMSE, Sharpe Ratio
- **Cloud et Big Data** : AWS, Google Cloud, Apache Spark
- **Soft Skills** : Leadership, communication claire, vulgarisation de concepts techniques, gestion du temps, travail d'équipe

## Formations

- **Master Data Spécialité Machine Learning** - CentraleSupélec (2023)
- **Licence STS Mathématiques** - ULCO Calais, Mention Bien (2021)
- **Bac S Spé Mathématiques** - Lycée Sophie-Berthelot, Calais (2018)

## Langues

- **Français** : Langue maternelle
- **Anglais** : Niveau B2 (Certifié CLES B2)

# Expériences Professionnelles

## Stage de fin d'études - Zoe12Tech (01 - 07 / 2023) – Startup Secteur Bancaire

### Mission 1 : Détection de fraudes bancaires

- **Situation** : L'entreprise souhaitait améliorer la détection des fraudes afin de protéger ses clients.
- **Tâche** : Créer un algorithme de détection des fraudes basé sur une classification non supervisée.
- **Action** : Mise en place de modèles tels que Isolation Forest, réseaux bayésiens, et modèles de graphes.  
Prétraitement des données pour éviter l'oversampling et l'undersampling.  
Évaluation via AUC-ROC, Recall, et F1-score.
- **Résultat** : Amélioration de la détection des fraudes avec une augmentation du recall de 15% tout en maintenant une interprétabilité optimale.

### Mission 2 : Prédiction de risque de crédit

- **Situation** : Besoin d'un modèle adaptatif pour évaluer le risque de crédit, aligné sur les objectifs des banques clientes.
- **Tâche** : Développer un modèle prédictif hybride intégrant des règles métiers modifiables.
- **Action** : Conception d'un modèle utilisant Random Forest et XGBoost, accompagné d'une surcouche de règles métiers paramétrables par les banques.  
Validation croisée et utilisation de SHAP pour garantir l'interprétabilité.
- **Résultat** : Modèle à la fois performant et adaptatif, évalué avec un AUC-ROC de 0,92.

## ALLA2 (Projet perso) (08-09 / 2023)

### Prédiction du prix du Bitcoin

- **Situation** : Intérêt personnel pour la Blockchain et la décentralisation.
- **Tâche** : Créer un modèle prédictif intégrant des métriques financières et des analyses de sentiment.
- **Action** : Enrichissement des données (SMA, RSI, MACD), analyse des sentiments du marché avec NLP et Google Trends.  
Modélisation avec GARCH, LSTM, et ensemble de modèles (stacking, bagging).
- **Résultat** : Prédiction précise avec un RMSE réduit de 20% par rapport à un benchmark simple.

## Missions Freelance

Olist (10-12 / 2023)



### Classification des clients selon leur comportement d'achat

- **Situation** : Comprendre les comportements d'achat pour améliorer les actions marketing.
- **Tâche** : Créer un modèle de clustering pour classer les clients.
- **Action** : Utilisation d'algorithmes non supervisés (Kmeans, DBSCAN), feature engineering avec PCA, et encadrement d'un apprenti pour la maintenance du modèle.
- **Résultat** : Réduction de 25% des coûts marketing grâce à une segmentation plus ciblée.

StackOverflow (01-03 / 2024)



### Modèle de suggestion de tags

- **Situation** : Améliorer la précision des suggestions de tags pour les questions posées sur StackOverflow.
- **Tâche** : Créer un modèle de classification multi-label pour prédire les tags pertinents.
- **Action** : Extraction de données via SQL, prétraitement du texte, représentation vectorielle (TF-IDF, Word2Vec, BERT), et classification multi-label avec logistic regression et transfert learning.
- **Résultat** : Augmentation de la précision des tags de 30%, améliorant l'expérience utilisateur.





### Reconnaissance de la Race de Chiens sur des Photos

- **Situation** : Besoin d'un modèle pour identifier les races de chiens afin de soutenir le personnel dans leurs missions.
- **Tâche** : Entraîner un modèle pour classer automatiquement les images de chiens et prédire leur race.
- **Action** : Utilisation de réseaux de neurones convolutifs (CNN) avec des modèles pré-entraînés (VGG16, ResNet, Inception, EfficientNet). Application de techniques de data augmentation (rotation, miroir, zoom) pour améliorer les performances.
- **Résultat** : Capacité à prédire les races les plus probables (>10%) pour aider à l'identification des potentiels croisements.



### Détection d'Anomalies en Temps Réel à partir de Vidéo

- **Situation** : Développer un système de surveillance pour détecter les anomalies dans les processus de production.
- **Tâche** : Mettre en place un modèle de détection d'anomalies à partir de vidéos en direct.
- **Action** : Utilisation d'algorithmes de détection d'objet (YOLO, SSD, Faster R-CNN) et d'algorithmes d'anomalie (Isolation Forest, Prophet). Intégration d'un système d'alertes en temps réel avec Apache Kafka et RabbitMQ.
- **Résultat** : Mise en place d'un système d'alerte performant permettant au personnel d'intervenir rapidement en cas d'anomalie détectée.

## Projets et Activités Secondaires

- **HermesWeb** : Suite de services pour petites entreprises, développement d'une présence en ligne solide, depuis janvier 2024. [www.hermes-web.com](http://www.hermes-web.com)
- **Owlearn** : Plateforme de mise en relation élève-professeur particulier (2021-2022)

## Centres d'intérêt

- **Karaté** : Champion départemental, Ceinture noire
- **Boxe Thaï** : Combat professionnel gagné par KO au premier round en Thaïlande
- **Marathons**

---

## Témoignages

"Thomas a su métamorphoser un problème complexe en une solution innovante et interprétable pour nos équipes non techniques. Son aptitude à vulgariser les concepts a été très appréciée."

– **Responsable Projet, Zoe12Tech**

"En tant que mentor, Thomas a fait preuve d'une pédagogie exceptionnelle, permettant à un junior de s'approprier les concepts avancés du machine learning tout en respectant les meilleures pratiques de clean code." – **Manager, Olist**