

Mouhamadou SECK

Data Scientist | AI Engineer | MLOps

✉ ahmadoubseck06@gmail.com

🏠 14 rue cité des platanes,
13005 Marseille

📍 France

📞 +33 6 09 14 67 57

Passionné par l'Intelligence Artificielle et son application concrète, j'ai développé une expertise en Machine Learning, Deep Learning, LLMs et analyse statistique grâce à mon parcours en Mathématiques Appliquées et Data Science à Aix-Marseille Université, enrichi par des expériences académiques et industrielles. Je souhaite mettre ces compétences au service de projets innovants en entreprise, en contribuant au développement de solutions d'IA performantes et à fort impact métier.

Diplômes et Formations

- **Master Mathématiques Appliquées et Statistiques – Data Science**
De septembre 2023 à septembre 2025 Aix Marseille Marseille, Saint Charles
- **Licence Métiers du Décisionnel et de la Statistique**
De septembre 2021 à septembre 2022 Aix Marseille Marseille, Luminy
- **Licence Mathématiques**
De septembre 2018 à juin 2021 Aix Marseille Marseille, Luminy

Expériences professionnelles

- **Stagiaire Data & IA – Développement de Chatbot RAG**
De mars 2025 à septembre 2025 ArcelorMittal, RDSF Fos-sur-Mer, France
 - Conception et implémentation d'un chatbot intelligent exploitant une architecture RAG (Retrieval-Augmented Generation) combinant LangChain, modèles open-source (Mistral, DeepSeek, Qwen, Gemma), embeddings avancés (BGE, E5, Nomic) et une base vectorielle locale (FAISS, Chroma).
 - Le système permet l'interrogation contextuelle de documents industriels (PDF, Markdown, DOCX, etc..) via une interface web moderne (Streamlit). Fonctionnalités clés : parsing automatique, chunking dynamique, génération de vecteurs, ranking par similarité cosine, requêtage sémantique, historisation des sessions et tuning interactif des hyperparamètres (température, max_tokens)
 - **Technos :** Python · LangChain · Streamlit · Hugging Face Hub · Transformers · LLMs open-source (Ollama) · Prompt Engineering · Marker · PyMuPDF · RAG pipeline

LSTM pour la Prédiction de la Vorticité – Projet académique M2 Data Science

De décembre 2024 à mars 2025

Projet Institut de Mathématiques de Marseille Marseille

- Lors d'un projet académique, j'ai conçu un modèle ConvLSTM2D pour anticiper l'évolution de la vorticité dans des fluides turbulents à partir d'images issues de simulations DNS. Pour améliorer l'efficacité, j'ai appliqué la transformée en ondelettes (PyWavelets) afin de compresser les données d'entrée, et optimisé les hyperparamètres pour un équilibre entre précision et performance. Le modèle a été entraîné sur un cluster HPC via Slurm, atteignant des résultats solides ($MSE = 0.0023$, $SSIM = 0.89$). Ce travail a été synthétisé dans un rapport scientifique complet en LaTeX.
- **Technos :** Python, TensorFlow, ConvLSTM2D, PyWavelets, HPC, Slurm, LaTeX

Alternant Data Scientist

De septembre 2023 à septembre 2024 ArcelorMittal Fos-Sur-Mer

- Dans le cadre d'un projet industriel majeur, j'ai analysé et quantifié l'origine des rayures sur les bandes d'acier lors du décapage, en utilisant notamment un modèle logistique. Grâce à des méthodes avancées d'analyse de données et à l'exploitation de données capteurs, j'ai identifié la vitesse minimale (V_{min}) comme facteur clé. Il a été démontré que 64,7% des rayures survenaient à cette vitesse, précisément au niveau des dômes séparant les bacs (1, 2 et 3). Ces travaux ont mené à l'élaboration d'une cartographie détaillée des causes, fournissant des solutions concrètes pour réduire ces défauts et optimiser la production.
- **Technos :** Python, Pandas, Scikit-learn, SQL, analyse de signaux, régression logistique, statistiques industrielles.



Langues

Français
Courant

Anglais
professionnel

École 42 Marseille (2025)

Linux / Shell
Commandes système, scripts automatisés, gestion processus & droits utilisateurs

Git
Versionning, branches, rebase, merge requests, travail collaboratif

Langage C
Pointeurs, allocation mémoire, gestion d'erreurs, fonctions & librairies

Méthodologie
Normes de code strictes (42 Norm), rigueur, clean code, peer-programming

Algorithmie
Résolution de problèmes, récursivité, parsing, manipulation de chaînes

Debugging
GDB, valgrind, gestion des fuites mémoire, segmentation fault

Atouts

Esprit analytique

Autonomie et rigueur

Adaptabilité

Esprit d'équipe

Curiosité

Apprentissage continu

Projets clés récents

- **Prédiction de la qualité de l'air avec un pipeline MLOps complet**
De juin 2025 à juillet 2025 Projet personnel Marseille, France
Développement d'un pipeline MLOps pour collecter automatiquement les données via API [aqicn.org](#), les transformer et les enrichir.
Entraînement d'un modèle **XGBoost** validé par découpage train/test.
Automatisation avec **GitLab CI/CD** et déploiement d'une app **Streamlit** pour visualisation en temps réel.
Respect des bonnes pratiques : modularité, environnements virtuels et versioning Git.
- **Analyse prédictive & inférence bayésienne**
De février 2025 à avril 2025 Aix-Marseille Université Marseille, France
Développement de modèles statistiques bayésiens (régression logistique et binomiale) pour prédire des défaillances et comparer des proportions. Utilisation de chaînes MCMC (20k itérations), diagnostics de convergence (\bar{R} , DIC) et interprétation via intervalles crédibles et visualisations a posteriori.
Technos : R, R2jags/JAGS, MCMC, modèles bayésiens.
- **Compressed Sensing – Fourier & Wavelets**
De novembre 2024 à janvier 2025 Aix-Marseille Université Marseille, France
Étude théorique et pratique de la reconstruction parcimonieuse de signaux et images à partir de mesures incomplètes. Implémentation d'algorithmes d'optimisation (ℓ_1 -minimisation, OMP, CoSaMP) et comparaison des approches Fourier vs Wavelets pour l'imagerie et le traitement du signal.
Technos : Python/R, optimisation convexe, traitement du signal, analyse spectrale.

Compétences

COMPÉTENCES TECHNIQUES ET SCIENTIFIQUES

- **Statistiques & Modélisation**
Solide maîtrise des méthodes statistiques classiques et avancées :
Modélisation prédictive : régression linéaire/logistique, arbres de décision, forêts aléatoires, XGBoost
Analyse exploratoire & multivariée : ACP, clustering, classification
Séries temporelles et modélisation bayésienne : processus ARMA, modèles probabilistes, MCMC, bootstrap
Compétence en optimisation de modèles : feature engineering, validation croisée, Bayesian Optimization
- **Machine Learning & Deep Learning**
Bonne maîtrise de l'apprentissage supervisé et non supervisé, incluant :
Autoencoders, détection d'anomalies
Conception, entraînement et mise en production de modèles profonds (CNN, Transformers)
Maîtrise des algorithmes EM, MCMC et méthodes variationnelles
Expertise en NLP avancé : LLMs, fine-tuning, modèles d'embedding, RAG, Ollama, Hugging Face, Mistral, LLAMA, DeepSeek
- **Programmation & Outils**
Expertise en Python : NumPy, Pandas, Scikit-Learn, TensorFlow, Keras, PyTorch, Streamlit
Bonne pratique en R (ggplot2, dplyr), SAS, Spark (PySpark)
Maîtrise des outils de versioning (Git, GitLab)
- **Bases de données & Big Data**
Maîtrise des bases SQL & NoSQL : PostgreSQL, MySQL, MongoDB
Connaissances du modèle Merise
Familiarité avec les environnements cloud et Big Data (Azure, Hadoop)
- **Visualisation & Aide à la décision**
Bonne compréhension des algorithmes d'optimisation (descente de gradient, Newton, KKT)
Visualisation et reporting : Power BI, Streamlit, Matplotlib, Seaborn, Plotly
Expérience en conception de tableaux de bord interactifs et visualisation de données complexes
- **Compétences théoriques**
Fondements de l'inférence statistique : estimateurs, méthodes paramétriques et non paramétriques, asymptotiques
Concepts clés en IA générative : embeddings, métriques de similarité, attention mechanism, architectures Transformers
- **Recherche & Communication scientifique**
Excellent rédaction scientifique (LaTeX), vulgarisation des résultats complexes
Aptitudes à la synthèse bibliographique, évaluation critique des méthodes, collaboration en recherche

Références

- **Freitas, Pedro Lucas**
Project leader Software Solutions, GSMPM, ArcelorMittal Méditerranée
pedro.lucasfreitas@arcelormittal.com
- **Dale, Franck**
Project Manager Software Solution Dvpt, ArcelorMittal Méditerranée
franck.dale@arcelormittal.com
- **Lionel DUFFO**
Ingénieur support process , ArcelorMittal Méditerranée
Lionel.duffo@arcelormittal.com

Expériences associatives

- **Mentor en mathématiques – Engagement bénévole – AFEV**
De septembre 2022 à juin 2024
Aix-Marseille Université Marseille, France
Encadrement personnalisé d'élèves du secondaire en difficulté, dans le cadre d'un programme associatif d'égalité des chances. Déconstruction des blocages mathématiques grâce à des explications claires, des exercices ciblés et un accompagnement régulier. Développement de méthodes d'apprentissage sur-mesure pour renforcer la compréhension, l'autonomie et la confiance en soi. Impact : amélioration significative des résultats scolaires (+2 à +4 points de moyenne) et regain d'implication en classe.

Centres d'intérêt

- Musculation, Football, Marche**
- Kaggle, Hackathons, Finance**
- Git Learn Application**
- Tech Programmation**

Réseaux sociaux

 @SeckMouhamadou