

**Amadou MAMANE**  
**Data Scientist / ML Engineer**  
**Disponible dans 3 mois**

✉ mamesark@gmail.com  
in https://fr.linkedin.com/in/amadoumamane  
g https://github.com/AmadouMamane

☎ +336.52.78.18.48  
🏠 Vincennes, France  
🌐 Français

## COMPÉTENCES

---

### ✓ Techniques

- Frameworks ML **Scikit-learn, Spark ML, TensorFlow, Pytorch, Vertex AI, Statsmodels**
- Visualisation de données **Matplotlib, Seaborn, Plotly, Bokeh, Tableau, Grafana**
- Langages de programmation **Python, Numpy, Pandas, Scipy, Scala, Java, Shell, SQL**
- Bases de données **Hive, BigQuery, Chroma, Pinecone, Cassandra, Oracle, OpenTSDB**
- MLOps **Git, GitLab CI/CD, Kubeflow, Airflow, Docker, Kubernetes**
- Plateformes big data **Hadoop, Spark, Flink, Kafka, Nifi, Streamsets**
- Cloud computing **Google Cloud Platform, AWS, Openstack**
- IA Générative **GPT, Llama, Gemini, Gemma, LangChain, LangGraph, HuggingFace**

### ✓ Fonctionnelles

- **Statistiques descriptives et inférentielles** (estimation, test d'hypothèse), **Séries temporelles**
- **NLP**, Théorie de graphes et analyse de réseaux, **Transformers, RAG, Prompt engineering**
- **Vision par ordinateur, Deep learning, Systèmes de recommandation, IA générative**
- Apprentissage non supervisé, Segmentation, **Méthodologie agile**
- Data engineering, **MLOps**

### ✓ Certifications

- C# developer
- Applied Data Science Program, MIT

### ✓ Langues

- **Français** Courant
- **Anglais** Professionnel

## FORMATION

---

**2024 - 2024** **Applied Data Science Program**: Utilisation de l'IA pour une prise de décision efficace, MIT

**2014 - 2015** **M.S. Ingénierie des Systèmes Informatiques Ouverts**, Ecole Centrale Paris

**2011 - 2012** Master recherche Génie Industriel et Informatique, **Ecole des Mines de Saint-Etienne**

## Data scientist

### Projets académiques et personnels

#### ✓ Détection du cancer de la peau (Compétition Kaggle ISIC-2024)

- Développement d'une solution de **détection du cancer de la peau** à l'aide d'**images** et de **métadonnées tabulaires**
- **Feature engineering** sur les données tabulaires
- Gestion du déséquilibre des classes :
  - **Augmentation** des images
  - **Recherche et intégration** de **données externes**
  - Application de **SMOTE** sur les données tabulaires
- **Fine tuning** de plusieurs **modèles de base**
  - VGG19
  - Resnet50
  - Resnet152V2
  - Densenet201
- Création de modèles avancés à l'aide de techniques de **validation croisée** et d'assemblage de modèles
- Implémentation/Adaptation de **fonctions de coût** et de **métriques de performance**
  - Focal loss, **weighted binary cross-entropy**, class balanced loss
  - **Partial AUC**, F1-score, recall
- Optimisation de la performance d'exécution sur CPU, **GPU** et TPU

#### ✓ Agent conversationnel interactif

- Création d'une application de commande dans un restaurant avec une interface de discussion en langage naturel permettant aux utilisateurs de passer commande
  - **Modélisation en structure de graphes** : actions, flux et boucles conversationnelles pour modéliser les étapes de la commande
  - Mise en place d'un **menu dynamique** et d'une **gestion d'état** pour suivre l'historique des commandes en temps réel
  - Ajout de **fonctions interactives** pour améliorer l'assistance client : ajustement de quantité/préférences, simulation de transmission en cuisine

#### ✓ Système de recommandation de musique

- Construction d'un système de **recommandation de musique** utilisant les données:
  - d'interaction utilisateur
  - de retour implicite (Implicit feedback)
  - des chansons
- Utilisation de techniques de :
  - Rang/popularité
  - **Clustering**
  - **Filtrage collaboratif**
  - **Filtrage collaboratif basé sur le contenu**
  - **Factorisation de matrice**
- Implémentation de la solution finale sous forme de **système de recommandation hybride** utilisant à la fois les méthodes de remplissage de matrice et de filtrage collaboratif à base de contenu

### Environnement technique

Google Colab, Scikit-learn, Python, Seaborn, Tensorflow, Surprise, LightFm, Nltk, Imblearn, LangGraph, API Gemini

### Environnement fonctionnel

Santé, Musique numérique, E-commerce

## EXPÉRIENCE PROFESSIONNELLE

Septembre 2023 – Novembre 2023

Cartes Bancaires, Paris

### Data scientist

*Au sein d'une équipe de 20 personnes*

- ✓ Développement d'une solution de **détection de fraudes** sur les commerçants en ligne à l'aide de données de type séries temporelles
  - Analyse exploratoire pour **identifier les comportements de fraude**
  - Modélisation de la fraude
  - **Labélisation des données** et **feature engineering**
  - Implémentation d'un modèle d'apprentissage supervisé à base de **gradient boosting**
- ✓ **Recommandations** de bonnes pratiques de développement **Apache Spark**

Environnement technique

PySpark, Spark ML, Python, Hive, Impala, Oozie, Git, Séries temporelles, Data leakage

Environnement fonctionnel

Banque, Lutte contre la fraude

Avril 2019 – Août 2023

RCI BANK&SERVICES, Paris

### Data scientist

*Au sein d'une équipe de 30 personnes*

- ✓ **Conception** et implémentation d'un **score de prédiction** de l'attrition des clients
- ✓ Développement d'une application de **déduplication de données clients**
- ✓ **Migration des applications** existantes depuis un cluster Hadoop on-premise vers le cloud (**Google Cloud Platform**)
- ✓ **Industrialisation des traitements** avec des pipelines CI/CD, Airflow et Kubeflow
- ✓ Conduite de plusieurs projets **preuves de concept (POCs)** pour évaluer et intégrer de nouvelles technologies

Environnement technique

Sckit-learn, Spark ML, Scala, Python, GraphFrames, GCP, Nifi, Gitlab CI/CD, Jira, NLP, Théorie de graphes

Environnement fonctionnel

Banque, Risque et crédit, Connaissance client, Méthodologie agile

Mai 2017 – Mars 2019

Altran/Orange, Arcueil Cachan

### ML engineer

*Au sein d'une équipe de 20 personnes :*

- ✓ Développement d'applications pour le **traitement en temps réel** des données 2G, 3G et 4G et mise à disposition de ces données dans un format pivot pour tout le groupe
- ✓ Participation au **développement, prototypage, test et sélection** de composants applicatifs
- ✓ Supervision de l'activité réseau **à l'aide KPIs** métiers
- ✓ **Mise en production** et **exploitation** de plusieurs applications

Environnement technique

Kafka, Flink, Spark, Hadoop, InfluxDB, Grafana, Streamsets, Scala, Java, Git, shell

Environnement fonctionnel

Télécommunications, Connaissance client

### Data scientist

Au sein d'une équipe de 20 personnes :

- ✓ **Collecte des données** web de recherches sur les trains
- ✓ Développement d'un **modèle** de prédiction de remplissage de la première classe
- ✓ Contribution à l'**amélioration de la qualité de code et mise en place de bonnes pratiques de développement**

Environnement technique

Scikit-learn, PySpark, Python, Pandas, Numpy, scala, Spark, Git, Bokeh, Jenkins, Luigi

Environnement fonctionnel

Yield management, Méthodologie agile

### Data scientist

Au sein d'une équipe de 30 personnes :

- ✓ Participation à l'implémentation de la solution **Customer Experience Management On Demand (CEMOD)** de Nokia permettant l'**amélioration de la performance du réseau mobile et de la connaissance client**
- ✓ Contribution à la mise en forme des données **réseaux, facturation, et CRM**
- ✓ Intégration des données dans le CEMOD et calcul de plusieurs **scores** pour évaluer la **satisfaction** client ainsi que la qualité du réseau dans les zones de vie du client.
- ✓ Participation à l'**anonymisation des données sensibles**

Environnement technique

Hive, Pig, Java, Oozie, shell, svn

Environnement fonctionnel

Télécommunications, Connaissance client

### Datascientist/Stage

Au sein d'une équipe de 20 personnes :

- ✓ Implémentation d'un **système de détection d'attaques par déni de service sur des logs webs.**
- ✓ Recherche sur l'état de l'art de l'analyse de logs webs
- ✓ Nettoyage des données et reconstruction des sessions de navigation
- ✓ **Feature engineering** et **visualisation des données**
- ✓ **Labélisation** des données et **clustering** pour identifier les groupes de sessions de **navigation**
- ✓ Développement et fine-tuning de plusieurs modèles de **classification** pour détecter les attaques par déni de service

Environnement technique

Hadoop, Scikit-learn, Pig, Hive, Python, Java, Pandas, Numpy, Scipy, Matplotlib, Git

Environnement fonctionnel

Banque, Méthodologie agile