

Lamine NDAO Data scientist Researcher (R&D)

(C) GitHub

Website

₽ PROFIL

Actuellement doctorant en Recherche et Innovation dans le domaine de la maintenance prédictive et de l'IA explicable (XAI), je suis ingénieur statisticien diplômé de l'ENSAE et titulaire d'un Master en Big Data & Data Science obtenu à Télécom Paris / Paris-Saclay.

Mon expertise couvre des domaines variés tels que la data management, la data science, le data mining, le traitement du langage naturel (NLP) et le MLOps. Je maîtrise parfaitement Python, R, ainsi que des outils et langages tels que SQL, PySpark et SAS. Mes différentes expériences m'ont permis d'acquérir une solide connaissance du secteur industriel, en particulier dans le cadre de la maintenance prédictive. Passionné par l'innovation, je m'intéresse particulièrement aux avancées dans le domaine de l'IA, notamment dans ses applications industrielles et sociétales.

E EPÉRIENCES PROFESSIONNELLES

LINEACT-CESI, Data scientist en R&D

En tant que doctorant, je participe à de différentes missions et activités liées à ma thématique de recherches.

2022 - aujourd'hui

Paris, France

Étude : Topic Modeling sur les publications autours de l'IA eXplicable (XAI).

- Collecte de publications sur l'IA eXplicable entre publiées entre 2015 et 2022 par Web scraping Python.
- Analyse des documents par une approche NLP permettant d'extraire les sujets traités par cette littérature (LDA).
- Outils techniques : Python, web scraping, Excel

Étude : Prédiction de la durée de vie RUL des moteurs du dataset C-MAPSS.

- Nettoyage des données de maintenance prédictive (séries temporelles)
- Mise en place d'un modèle de type LSTM capable de prendre en compte l'aspect séguentielle des données.
- Prédiction et Explication via les méthodes LIME, SHAP et L2X.
- Outils techniques : Python (tf), shap, lime, git

Chargé de cours et de Workshop en option FISA

- Traitement de données et réseaux de neurones : Initiation à Tensorflow
- Reseaux de Neurones Convolutifs (CNNs)
- Traitement du langage naturel et classification par Réseaux de Neurones Récurrents (RNNs)
- Auto-encodeur Convolutif et traitement d'images

CNAM-CEDRIC

Chargé de cours / TD du module STA101 - Présentiel

Analyse des données : méthodes descriptives

- Analyse bivariée

- Algèbre linéaire 1, 2 and 3
- Analyse Multidimensionnelle (ACP, AFC, ACM, AFDM)
- Classification (kmeans, CAH)
- Programmation R, Python, SAS
- Outils : R, Python, SAS, Excel

2023 - aujourd'hui Paris, France

Chargé de cours / TD du module Entreposage et fouille de données (STA211) Réseaux de neurones et apprentissage profond

2023 - aujourd'hui Paris, France

- Modèles prévisionnels et systèmes de gestion de l'entreprise
- Méthodologies de Data Mining
- Pré-traitement des données
- Données et techniques de fouille
- Méthodes non supervisées
- **Outils** : Environnements freeware : R, Python- Outils spécifiques : SAS-EM, SPAD- Data Mining et bases de données : OLAP Business Object

Ingénieur de Recherches

Mise en place d'une méthode pour la gestion de la confidentialité des données en apprentissage distribué à partir d'objets connectés :

- Études théoriques via une revue de la littérature
- Mise en place d'un plan d'expérimentation
- Implémentation de la méthode sous Python

Résultats : À partir des données distribuées, la méthode implémentée "e-Privacy" a permis de ne pas divulguer les informations à caractère individuel.

Outils techniques: Python, PySpark, GCP, SQL

Data Scientist

Consensus pondéré pour la gestion des données manquantes en classification non supervisée.

- Mise en place d'un algorithme de Clustering capable d'apprendre sur des données multi-tableaux (issues de différentes sources sur une même population) non complète.
- Ensemble Learning, Classification multi-blocs, consensus de partitions pondéré et non pondéré, Méthodes d'imputation multiple.
- Développement de packages sur R, Python.

Outils techniques: R, Python, SAS, SQL

STATINFO, Data Analyst 🖸

Cabinet d'expertise spécialisé dans la conception et le management d'enquête, la formation dans les outils de traitement de données.

Rôles et responsabilités :

- Traitement de données d'enquête et Rédaction de rapports d'études.
- Former des agents de l'État dans les outils de traitement de données à travers le PAPUSGS.
- Réalisation de tableaux de bord de suivi des API liés à la qualité des données collectées

Outils techniques :

- R, SPSS, Python, Tableau, MS Excel VBA, ONA, ODK collect

IRD, UMR VITROME, Statisticien | Chargé d'études ☑

Utiliser les outils statistiques pour évaluer quantitativement "La confiance institutionnelle en Afrique, interactions avec le système de soins, opinions sur les politiques de santé et comparaisons internationales."

- Mission :

Exploiter les données d'enquête d'afro-baromètre pour mettre en place un score permettant de comparer le niveau de confiance en les institutions des pays africains.

Taches:

- Dépouillement et prétraitement des données d'enquête à l'aide de R et SPSS.
- Analyse Exploratoire des données d'enquête à l'aide de R et Python.
- Mise en place d'un score numérique en utilisant les outils d'analyse factorielle comme ACM.
- Représentation cartographique à l'aide de l'outil QGIS

09/2021 - 09/2022

Paris, France

04/2021 - 09/2021

Paris, France

07/2020 - 09/2020

Dakar, Sénégal

02/2020 - 06/2020

Dakar, Sénégal

- Mise en place d'un modèle explicatif du score par un modèle linéaire multiniveaux.

Outils techniques : R, Python, SPSS, QGIS

ANSD, Statisticien | chargé d'études 🗹

08/2018 - 10/2018

Mission : Analyse statistique descriptive, rédaction de rapports mensuels sur les Indices des Prix à la Production (IPP).

Dakar, Sénégal

Taches:

- Analyse descriptive des données temporelles
- Rédaction et publication de rapports mensuels
- Calcul d'indicateurs statistiques sur les IPP

Outils: Excel VBA, SPSS, R

FORMATIONS

Thèse en R&D, Conservatoire National des Arts et Métiers

10/2022 - aujourd'hui

Sujet : Vers une intelligence artificielle eXplicable (XAI) et équitable

Paris, France

Master 2, Université de Paris Saclay / Télécom Paris

09/2020 - 09/2021

Traitement de l'Information et Exploitation des Données (TRIED)

Paris, France

- Machine Learning, Deep Learning, Traitement d'images, Big Data, Statistique, Informatique, Gestion de base de données.

- Mention : Bien

Ingénieur Statisticien, ENASE

09/2016 - 06/2020

- Analyse Statistique, Informatique décisionnelle, Économétrie

Dakar, Sénégal

- Mention : Bien

A COMPÉTENCES TECHNIQUES

- Python (TensorFlow, Keras, Plotly, Scikit-learn, Numpy, Pandas, Librosa)
- Spark, SQL, R, Git, Neo4j, cypher, LLMs, Langchain
- SQL SAS (Data step, Proc SQL, Macro SAS)
- PowerBI, Tableau, Talend
- Hugging face, BERT, GPT

Intelligence Artificielle

Collaboration

Deep Learning, Machine Learning, Natural Language Processing, Audio processing, Recommendation Systems

Git, Google colab, Méthodologie agile

DataViz, analyses descriptives, reportings,

Data Analyse

IA Generative

Multiplex Networks, Graph databases: Neo4j, Cypher query Language, LLMs.

• • • • contrôle de qualité, Analyse des besoins métier

Technologies & Langage de

Programmation

Python (TensorFlow, Keras, Plotly, Scikitlearn, Numpy, Pandas, Librosa), Spark, SQL, R, Git, Neo4j, cypher, Langchain, SQL - SAS (Data step, Proc SQL, Macro SAS)

COMPÉTENCES FONCTIONNELLES

Maintenance prédictive

Détection de fraude

Détection d'anomalies, Prédiction du RUL

Ciblage & connaissance client

Modélisation supervisée et non supervisée

Marketing

Collecte de données

Traitement d'images et des signaux

Outils de collecte : ODK collect, ona,

Commcare.

Web Scraping (Python : BeautifulSoup,

requests)

(6) LANGUES

Anglais Français

Bright English (B2)

© CERTIFICATIONS

- Introduction to Machine Learning in Production (Coursera) 🗹

- Spark avec Python -Pratique avec le Big Data (Udemy) 🗹

- Getting Started with AWS Machine Learning (Coursera) 🗹

- Google Cloud Big Data and Machine Learning Fundamentals (Coursera) 🗹

15/03/2025

2024

2023

□ PUBLICATIONS

Improving predictive maintenance: Evaluating the impact of preprocessing and model complexity on the effectiveness of eXplainable Artificial Intelligence methods, Engineering Applications of Artificial Intelligence, Pergamon

Effet de la complexité du réseau LSTM sur l'explicabilité en Maintenance 2024 Prédictive, JdS 2024 : 55ièmes Journées de Statistique 🗹

Enhancing Explainability in Predictive Maintenance: Investigating the Impact of Data Preprocessing Techniques on XAI Effectiveness, The 37th International Conference of the Florida Artificial Intelligence Research Society-FLAIRS-37 🛮

Consensus de partitions en NLP pour une revue systématique de la littérature 2023 autour de l'XAI du biais et de l'équité, SFC'2023

Systematic review of the literature on bias, fairness and explicability (XAI),

CNIA'2023 🗹

ATOUTS

Rigueur - Confirmé **Autonomie** - Confirmé

adaptation - Confirmé Organisation - Confirmé

Esprit d'analyse - Confirmé

⟨of CENTRE D'INTÉRÊTS

Sport - Jogging, fitness | Lecture - Innovation, Economie, technologie | Veille technologique (IA)