

Lamine NDAO

Data scientist Researcher (R&D)



✉ mlndao@cesi.fr

☎ 0627487763

📍 Paris, France

in www.linkedin.com/in/mlndao

🔗 <https://github.com/laminendao>

👉 flowcv.me/lamine-ndao-nd

LANGUES

Anglais ● ● ● ● ●
Bright English (B2)

Français ● ● ● ● ●

COMPÉTENCES FONCTIONNELLES

Maintenance prédictive
(Détection d'anomalies, Prédiction du RUL)

Détection de fraude

Modélisation supervisée et non supervisée

Ciblage & connaissance client
(Marketing)

Collecte de données (Outils de collecte : ODK collect, ona, Commcare. Web Scraping (Python : BeautifulSoup, requests))

Traitement d'images et des signaux

PROFILE

Actuellement en Doctorat de R&D en Data dans le domaine de la Maintenance prédictive, je suis Ingénieur statisticien (ENSAE) et titulaire d'un Master en Big Data & Data Science, obtenu à Télécom Paris / Paris Saclay. J'interviens sur des problématiques de data management, data science, data mining, NLP et MLOps. Je dispose d'une très bonne maîtrise de Python, R et des logiciels et des langages SQL, PySpark et SAS. À travers mes différentes expériences, je dispose également d'une bonne connaissance du secteur Industrielle dans la Maintenance Prédictive.

EXPÉRIENCES PROFESSIONNELLES

Data scientist en R&D, LINEACT-CESI

2022 – present | Paris, France

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

Étude : Topic Modeling sur les publications autour 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 séquentielle 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
- Réseaux de Neurones Convolutifs (CNNs)
- Traitement du langage naturel et classification par Réseaux de Neurones Récurrents (RNNs)
- Auto-encodeur Convolutif et traitement d'images

Chargé de cours/TD, CNAM

2023 – present | Paris, France

Chargé de 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

Ingénieur de Recherches, CNAM-CEDRIC

09/2021 – 09/2022 | Paris, France

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.

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

PUBLICATIONS

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

2023

Systematic review of the literature on bias, fairness and explicability (XAI), CNIA'2023

2023

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)

Outils techniques : Python, PySpark, GCP, SQL

Data Scientist, CNAM-MSDMA

04/2021 – 09/2021 | Paris, France

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

Data Analyst, STATINFO

07/2020 – 09/2020 | Dakar, Sénégal

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

Statisticien | Chargé d'études, IRD, UMR VITROME

02/2020 – 06/2020 | Dakar, Sénégal

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

Outils techniques : R, Python, SPSS, QGIS

Statisticien | chargé d'études, ANSD

08/2018 – 10/2018 | Dakar, Sénégal

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

Taches :

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

CENTRE D'INTÉRÊTS

- Sport
- Lecture
- Veille technologique (IA)

ATOUTS

Rigueur	● ● ● ● ●
Autonomie	● ● ● ● ●
adaptation	● ● ● ● ●
Organisation	● ● ● ● ●
Esprit d'analyse	● ● ● ● ●

Outils : Excel VBA, SPSS, R

FORMATIONS

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

10/2022 – [present](#) | Paris, France

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

Master 2, Université de Paris Saclay / Télécom Paris

09/2020 – 09/2021 | Paris, France

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

- 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 | Dakar, Sénégal

- Analyse Statistique, Informatique décisionnelle, Économétrie
- Mention : **Bien**

COMPÉTENCES

Intelligence Artificielle

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

Collaboration

Git, Google colab, Méthodologie agile

IA Generative

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

Data Analyse

DataViz, analyses descriptives, reportings, contrôle de qualité, Analyse des besoins métier

Technologies & Langage de Programmation

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