

DOUNIA MOUCHRIF

DATA SCIENCE | IA | MODÉLISATION STATISTIQUE

douniamouchrif@outlook.com | +33 6 52 42 64 29 | [Linkedin](#) | [Portfolio](#)

Diplômée d'un Master en Data Science et disponible immédiatement, je suis à la recherche d'un poste me permettant de relever de nouveaux défis et de transformer les données en décisions et en valeur concrète.

FORMATION

Master - CMI Ingénierie de la Statistique et de l'Informatique

09/2023 - 09/2025

Université de Bordeaux

Analyse Classification Indexation des données, Modèles de régression, Modèles linéaires mixtes et applications en bio- statistique, Bases de données avancées, Probabilités et Statistiques, Apprentissage supervisé et non supervisé, Apprentissage profond, Sciences des données et des réseaux

Programme Erasmus - Informatique

01/2023 - 07/2023

Université Libre de Bruxelles

Licence - CMI Ingénierie de la Statistique et de l'Informatique

09/2020 - 07/2023

Université de Bordeaux

COMPÉTENCES

Programmation : Python (Pandas, Scikit-learn, PyTorch, TensorFlow, Matplotlib, Plotly, Seaborn), R, SQL

Outils : PostgreSQL, MongoDB, Git/GitHub, Jupyter Notebook - notions : Dataiku, Docker, FastAPI, MLflow

Data Science : Prétraitement et analyse statistiques des données, Machine Learning, Deep Learning, Visualisation de données et Applications Data (Streamlit, Dash, Shiny)

Langues : Français (langue maternelle), Anglais (B2), Espagnol (notions)

Transversales : Esprit d'équipe, Autonomie, Capacité d'apprentissage, Vulgarisation

EXPÉRIENCE

Stage en Data Science, Topnir Systems

04/2025 - 09/2025

- Étude de méthodes d'apprentissage automatique pour la prédiction des propriétés physico-chimiques de produits pétroliers à partir de données spectrales.
- Exploration et comparaison de plusieurs approches IA : détection d'outliers, réduction de dimension, méthodes de projection, clustering, classification, modèles de prédiction (Random Forest, XGBoost, LightGBM) et réseaux de neurones.
- Évaluation des performances des modèles et comparaison avec les solutions internes existantes
- Participation à la rédaction d'un article interne présentant les résultats et conclusions de l'étude

PROJETS

Pipeline ML de prédiction du churn client (projet personnel)

- Construction d'un pipeline ML sur le dataset Telco Customer Churn : preprocessing, feature engineering, optimisation d'hyperparamètres (Optuna) et suivi des expériences (MLflow).
- Déploiement du modèle via une API FastAPI et dashboard Streamlit, conteneurisés avec Docker.

Datathon Bordeaux Métropole - 48h pour transformer les données en indicateurs clés des transitions économiques

- Analyse des inégalités salariales et des organisations de l'économie sociale.
- Équipe gagnante du défi sur les transitions sociales et prix "Coup de coeur" du public.

Défi IA - Carrefour - Prédiction du réachat de produits

- Prétraitement des données, feature engineering et entraînement de modèles LGBM et XGBoost.
- Augmentation du Hit Rate@10 de 5 % par rapport à la solution initiale.

Classification automatique d'images satellitaires - Détection d'éoliennes

- Prétraitement des images, entraînement et optimisation de modèles.
- Conception de CNNs et transfer learning.