

# Linh Nguyen

✉ [linhnguyenlethuy1611@gmail.com](mailto:linhnguyenlethuy1611@gmail.com) ☎ +33767859415 📍 France 🔗 <https://www.linkedin.com/in/le-thuy-linh-nguyen/>  
🌐 <https://github.com/linhnguyenlethuy1611>

## PROFILE

Data Scientist polyvalente, maîtrisant modélisation prédictive, visualisation et analyse multi-sectorielle. Capable d'adapter mes compétences à divers contextes pour transformer des données complexes en solutions concrètes. Orientée collaboration et évolution vers des responsabilités élargies.

## EXPÉRIENCE PROFESSIONNELLE

### Data Scientist - OP FROM NORD – Boulogne-sur-Mer, Hauts-de-France, France

Jun 2024 - Sept 2025

- Déploiement d'une méthodologie d'analyse des données VMS (Vessel Monitoring System) avec Python et scikit-learn, automatisant des algorithmes de détection de trajectoires et réduisant le temps d'analyse de 80 %.
- Automatisation de processus de classification et de gestion des flottilles via APIs, SQL et pipelines de données, afin d'améliorer le suivi d'activité et la coordination des navires.
- Croisement de bases de données hétérogènes (logbooks, ventes, VMS), avec des techniques avancées de data wrangling, pour attribuer les captures et le chiffre d'affaires aux positions VMS avec une précision de 95 %.
- Réalisation d'analyses statistiques (régression, analyse exploratoire) pour étudier l'impact de la pêche selon les engins et zones géographiques, et création de visualisations via Tableau et Power BI pour appuyer la gestion des quotas.
- Développement de scripts de web scraping automatisés avec Selenium pour enrichir la base de données interne.
- Formation interne : animation de sessions de formation en Python et bonnes pratiques data science pour les collègues non techniques, favorisant la montée en compétence collective et la transversalité des savoirs.

### Data Analyst Environnementale - CONINCO., JSC – Hanoï, Vietnam

2018 - 2021

- Analyse de données environnementales avec **Python et QGIS** pour identifier les impacts écologiques.
- Création de **cartographies interactives** et **dashboards** via **Tableau / Power BI** pour faciliter la prise de décision.
- Rédaction de rapports clairs et communication avec les parties prenantes pour soutenir des stratégies durables.

### Analyste de données et Exécuteur modèle (Stagiaire master) - Laboratoire d'Océanologie et de Géosciences, Wimereux, France

Mars 2018 - Sept 2018

- Réalisation d'une modélisation écologique prédictive avec Python, à partir de données climatiques issues des scénarios du GIEC, pour anticiper la distribution future des espèces de poissons commerciaux.
- Analyse géospatiale des habitats marins à l'aide de QGIS et de systèmes d'information géographique (SIG), afin d'évaluer l'impact du changement climatique sur la biodiversité marine.
- Contribution aux stratégies de développement durable à travers la production de rapports scientifiques orientés vers la gestion raisonnée des ressources halieutiques et la préservation des écosystèmes.

## COMPÉTENCES

- |   |   |
|---|---|
| • Langages & Programmation: Python, R, SQL                                  | • Web Scraping & Automatisation : BeautifulSoup, Selenium                             |
| • Machine Learning & Deep Learning : scikit-learn, TensorFlow, PyTorch, NLP | • Big Data & Cloud : Hadoop, Spark, Azure, AWS, APIs                                  |
| • Visualisation & BI : Tableau, Power BI                                    | • Statistiques : statistiques avancées, modélisation prédictive, analyse exploratoire |
| • Analyse géospatiale : QGIS, ArcGIS  |   |

## PROJETS

### Projet Data Science – Retail Analytics (2025)

- Réalisé une **segmentation clients** à partir des données transactionnelles (Frequency, Recency, Basket size, Diversity) via K-Means, pour identifier des groupes exploitables (loyaux, à risque, acheteurs occasionnels).
- Développé un modèle de **recommandation de produits** (analyse de paniers fréquents et séquences d'achats) pour anticiper le prochain achat et personnaliser les offres.
- Construit un modèle de **prévision des ventes** (séries temporelles) pour estimer le volume de transactions hebdomadaires et optimiser la gestion des stocks.

### Développer un algorithme, à partir des données collectées, pour prédire la distribution des espèces de la famille des Scombridés (2024)

- Créer un jeu de données d'entraînement, collecter les données depuis deux plateformes et les apparier selon leurs coordonnées et leur date (Python)
- Développer un algorithme en utilisant une Machine à Vecteurs de Support (SVM)
- Créer une carte prédisant la probabilité d'apparition des Scombridés le long des côtes mondiales

### Analyste de Données Environnementales (2023)

- Analyse des tendances l'Indice de Qualité de l'Air (AQI)US (1980–2022) avec Pandas et modèles temporels.
- Visualisations interactives des 5 États les plus à risque (Matplotlib/Seaborn).
- Étude multivariée des polluants clés (O<sub>3</sub>, CO, NO<sub>2</sub>, PM2.5, PM10) et de leur impact sur l'AQI.
- Dashboard interactif de restitution (Dash/Plotly) pour publics non techniques.

## FORMATION

### Master de Sciences Technologies Sante Institut national polytechnique de Toulouse, France (2016-2018)

### Maîtriser l'analyse des données avec Pandas IBM - Coursera, 2022

### Certificat professionnel d'analyste de données IBM - Coursera, 2023

## LANGUAGES

- |  |                           |
|--|---------------------------|
| • Français: compétence professionnelle avancée | • Anglais : niveau avancé |
|--|---------------------------|