

Komlan D. KATAKOU

Statisticien / Data scientist



- doigtekatakou@gmail.com
- 0753936172
- 127 Rue Louis Becker, Villeurbanne (69100)
- <https://www.linkedin.com/in/komlan-doigte-katakou/>
- <https://github.com/KomlanDoigteKatakou>

COMPÉTENCES

- Modélisation statistique, Tests d'hypothèses et Prise de décision
- Maîtrise des modèles de Machine Learning et de Deep Learning avec implémentation sous Scikit Learn et Keras
- Data visualisation
- Création de Dashboards (Shiny, Power BI)
- Python

R

SAS

SQL

Git

LANGUES

- Français
- Anglais
- Espagnol

ATOUTS

- Ponctualité
- Créativité
- Facilité d'adaptation
- Bon relationnel

CENTRES D'INTÉRÊT

- Guitare
- Scrabble, Echecs
- Poésie (Pléiade, Parnasse, Surréalisme)

MON PROFIL

Passionné de statistique et d'apprentissage automatique, je recherche un stage ou une alternance en **Statistique / Data Science/ Machine Learning** pour mettre en oeuvre ma rigueur statistique, mes compétences en programmation et mon expérience en manipulation de données réelles et complexes.

MON CURSUS ACTUEL

**Master Statistique**, Université Claude Bernard Lyon 1  
09/2024 – aujourd'hui | Villeurbanne, France  
**\*\*\*Moyenne semestrielle : 16.54 / 20 (2ème sur 30)\*\*\***

MON EXPÉRIENCE

**Tutorat Sciences - Université Lyon 1**, Tuteur  
09/2024 – aujourd'hui | Lyon, France  
Accompagnement d'étudiants de niveau Licence pour la compréhension de cours et la résolution d'exercices

**Auchan Supermarché, Lyon - La Duchère**,  
Hôte de caisse et chargé du rayon fruits et légumes  
12/2023 – aujourd'hui | Lyon, France

**Institut National de la Statistique et des Etudes Economiques et Démographiques**, Stagiaire : Conception d'un modèle économétrique pour l'estimation du PIB régional (Logiciels R et Excel)  
06/2023 – 08/2023 | Lomé, Togo

- Collecte, nettoyage et structuration de bases de données régionales multi-sources (emploi, production, fiscalité, consommation, etc.)
- Analyse exploratoire et tests d'hypothèses
- Conception, estimation et validation de modèles économétriques (régression multiple, modèles à effets fixes, ACP + régression)

MES PROJETS

● **Prévision de la production éolienne chez RTE**,  
RTE et Université Lyon 1  
01/2025 – 04/2025  
Ce projet vise à modéliser la production éolienne française et à prédire le rendement du parc éolien. Nous disposons d'un modèle initial proposé par RTE, avec une erreur de prédiction moyenne de 4.7% sur le rendement et notre but est de proposer des modèles plus performants afin de réduire cette erreur et obtenir des prédictions plus précises.

- ♦ **Analyse cartographique et sectorisation climatique :**
  - Manipulation des coordonnées géographiques (longitude et latitude) des sites d'implantation d'éoliennes et visualisation cartographique
  - Découpage du territoire français en dix zones, en tenant compte des disparités climatiques régionales.
- ♦ **Traitement des données massives de RTE :** Extraction, Lecture, Nettoyage et Agrégation des séries chronologiques de données :
  - Données climatiques par site : Vitesses (composantes U et V) du vent à 100 mètres d'altitude, enregistrées chaque heure (2017 - 2022)
  - Données de production éolienne (mesurées chaque demi-heure sur la même période) à l'échelle nationale
- ♦ **Manipulation avancée du logiciel R :**
  - Utilisation intensive de packages variés : tidyverse, data.table, lubridate, forecast, maps, ggplot2, ...
  - Gestion de divers formats de fichiers (csv, xls, nc, RDS, dat, txt, ...)
- ♦ **Modélisation prédictive :**
  - Modélisation GAM : un premier modèle pour expliquer la puissance installée du parc éolien en fonction des variables climatiques et de la saisonnalité et un deuxième pour prédire le facteur de charge (indice de rendement) éolien à partir de la vitesse du vent sur les dix zones
  - Détection de multicollinéarité et amélioration des performances par ACP et des modèles de régularisation

**Résultat obtenu :** une erreur de prédiction moyenne (RMSE) de 0.04% sur le rendement, soit 100 fois mieux que le modèle initial de RTE  
**Autre impact :** Mise en évidence des variables les plus influentes sur le rendement éolien; Identification des zones à grand rendement et proposition de délocalisation des installations peu rentables.

Voir plus sur mon Github : <https://github.com/KomlanDoigteKatakou>