

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 (Excel, Seaborn, ggplot) et Création de Dashboards interactifs (R-Shiny, Dash, Power BI)

Gestion de bases de données : PostgreSQL, MySQL

LANGUES

- Français
- Anglais
- Espagnol

ATOUTS

- Ponctualité
- Créativité
- Facilité d'adaptation

CENTRES D'INTÉRÊT

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

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**, Hôte de caisse  
12/2023 – aujourd'hui | Lyon, France

MES PROJETS

- **Prévision de la demande en produits sanguins**

L'approvisionnement en produits sanguins est crucial : mal anticipé, il cause pénuries vitales ou surstocks coûteux. Ce projet, basé sur les données de l'Etablissement Français du Sang, vise à optimiser l'approvisionnement des hôpitaux via un outil d'aide à la décision.

- ♦ **Extraction et nettoyage des données** (Python, Pandas, SQL)
  - ♦ **Analyse statistique de la Demande :**
    - Visualisation avancée (avec Seaborn) pour détecter les tendances et notamment la cyclicité des demandes
    - Tests statistiques pour valider les hypothèses business (par exemple, lien entre météo et dons).
  - ♦ **Modélisation prédictive :** Construire un modèle robuste pour anticiper la demande et éviter les pénuries.
  - **Approche statistique :** Mise en œuvre d'un modèle de séries temporelles SARIMA pour capter les tendances saisonnières
  - **Approche Machine Learning :** Modèle Random Forest pour pallier les insuffisances du modèle SARIMA et améliorer les prévisions
  - **Approche Deep Learning :** Modèle LSTM pour capturer les schémas temporels complexes et expliquer la variabilité des demandes
  - ♦ **Déploiement et Automatisation :** Transformer le modèle en solution concrète pour les décideurs

Création d'un dashboard interactif (Power BI) pour suivre en temps réel les prévisions de demande et les alertes de pénurie. Mise à jour continue des prévisions via FastAPI pour automatiser les décisions.

  - ♦ **Impact Business du projet :**
    - Réduction de 20% des pénuries de sang grâce à une meilleure anticipation et Diminution du gaspillage de 15-30%
    - Gain de temps pour les décideurs, en remplaçant les décisions manuelles par un modèle data-driven.

● **Prévision de la production éolienne chez RTE**

Ce projet vise à analyser et prédire la production éolienne nationale à partir des données météorologiques agrégées sur 10 zones, en déployant des techniques de modélisation statistique avancée

- ♦ **Manipulation des données cartographiques** des sites d'implantation d'éoliennes et Découpage de la France en une grille de 10 zones en tenant compte des disparités climatiques régionales
  - ♦ **Traitement des données massives de RTE :** Extraction, Lecture, Nettoyage et Agrégation des données de production éolienne et des données météorologiques (données horaires sur dix ans)
  - ♦ Manipulation de packages variés du logiciel R : **dplyr, data.table, tidyverse, purrr, lubridate, forecast, maps, ggplot2**, ... et de fichiers à divers formats (**csv, xls, netcdf, RDS, dat, txt**, ...)
  - ♦ **Modélisation statistique :** Modèles **GLM** et **GAM** pour prédire la quantité de production éolienne en fonction des variables climatiques
  - ♦ **Développement d'un tableau de bord interactif** sous R-Shiny pour visualiser les prévisions en temps réel et guider la prise de décision en termes de politique énergétique par région.