

NGO Khoa Anh

khoa.ngo@etu.sorbonne-universite.fr | +33 663186491 | [linkedin/khoa-anh-ngo-5985a9156/](https://www.linkedin.com/company/khoa-anh-ngo-5985a9156/) | [github/ngokhoaanh](https://github.com/ngokhoaanh)

EDUCATION

Master 2 ISUP - Ingénierie Statistique et Data Science

Paris, FR | 2022-2023

SORBONNE UNIVERSITÉ

Cours: Machine Learning, Réseaux neuronaux, Apprentissage statistique, Séries chronologiques, Statistiques industrielles, Apprentissage pour la prévision, Calcul stochastique, Optimisation convexe en ligne, Modèle aléatoire, Méthodes de simulation pour les modèles génératifs

Master 1 ISIFAR-Ingénierie Statistique et Informatique, de la Finance, de l'Assurance et du Risque Paris, FR | 2021-2022

UNIVERSITÉ DE PARIS

Cours: Probabilités et Extrêmes, Programmation C, Analyse de Données, Actuariat, Mathématiques Financières, Base de Données avancées, Statistiques

Licence de Mathématiques

Paris, FR | Septembre 2018-Septembre 2021

SORBONNE UNIVERSITÉ

EXPÉRIENCE

COVÉA | CHARGÉ D'ÉTUDES, GESTION ET PILOTAGE

Paris, FR | 2022-2023

- Production des différents reporting en prenant en compte les évolutions produits et réglementaires.
- Mise en place de contrôles de conformité.
- Intégrer le programme de transformation DATA de Covéa.
- Participer aux évolutions des outils et tableaux de bords, traiter sous **SAS VA**, **Excel** et **VBA**

PROJETS

CLASSIFICATION SUR LE DIABETE

PYTHON, SCIKIT-LEARN, TENSORFLOW, SCIPY, KERAS

- Concevoir un modèle d'apprentissage permettant de prédire avec le moins d'erreur possible.
- Identifier des possibles réduction de dimension avec ACP.
- Regrouper les diabétiques et non-diabétiques par K-MEANS.
- Agrégation des différents modèles (Decision tree, Regression logistique, Random forest, SVM, Gradient boosting, lightGBM,...) et aussi des Reseaux Neuronaux pour donner un modèle final plus performant.
- Oversampling pour résoudre le problème de classes déséquilibrées.

WIKIPEDIA TRAFFIC FORECAST

PYTHON, SCIKIT-LEARN, TENSORFLOW, TORCH, STATSMODEL

- Comprendre comment les utilisateurs interagissent avec la plateforme et pour optimiser les performances et la disponibilité des serveurs de Wikipédia.
- Vérifier la stationnarité d'une série chronologique.
- Prédiction par modèles mathématiques (ARIMA, XGboost, Random forest) avec Python et Scikit-learn.
- Prédiction par Deep learning avec des réseaux récurrents sous Tensorflow.

COMPÉTENCES

Programmation: Python, C++, SQL, R, Tableau, SAS, CUDA, VBA

Librairies: Scikit-learn, Tensorflow, Keras, Pandas, Matplotlib, Numpy...

Langues: Français(C1), Anglais(B2), Vietnamien(Natif)

Misc: Git, Github, \LaTeX