

Minh Anh Nguyen

📞 Numéro de téléphone: (+33) 0695101063 ✉ Adresse électronique: manhnguyen1161998@gmail.com

🌐 Site web: <https://www.linkedin.com/in/minh-anh-nguyen-6b2009214/>

📍 Domicile: 12 Rue Henri Brisson, 78500 SARTROUVILLE (France)

EXPÉRIENCE PROFESSIONNELLE

Ingénieur data scientist - Alternance

Stellantis [09/2021 – 09/2023]

Localité: Poissy

Pays: France

Diagnostic et Fiabilité du chaine de traction

Développer des méthodologies et un ensemble d'algorithmes permettant de modéliser les queues de distribution de probabilités d'évènements rares, afin d'estimer le risque d'observation de valeurs extrêmes.

- Analyse statistique des données de roulages
- Examiner les modèles, les tendances et les caractéristiques des données collectées sur les véhicules dans des conditions normales d'exploitation.

- Amélioration la robustesse du diagnostic

- Etat de l'art et benchmark de méthodologies

- Développement des méthodologies et outils numériques.

- Data visualization Environnement technique: **Python, R**

Libraries: **scikit-learn, scipy.stats, numpy, pandas, multiprocessing, matplotlib.pyplot, scipy.interpolate, scipy.optimize, scipy.integrate**

R packages: climod, kdensity, reshape2, rtop, stats

PROJETS

Data Analytic et Data Lake: Impact de la Communication Scientifique sur Twitter

[11/2022 – 02/2023]

Exécution de processus **ETL**:

- Concevoir l'architecture, extraire de la source des APIs de Twitter en utilisant leur API pour récupérer données (OAuth 2.0, authentification, token etc.) : Tweepy

- Convertir des données en csv et json en parquets avant les uploader vers data lake : Pandas Nettoyer, Structurer, Filtrer les données en utilisant **pandas, Spark SQL**

Création de modèles de régression et de tests statistiques sous **R**.

Analyser de données en utilisant **Pyspark, Spark SQL, numpy, nltk** SentimentIntensityAnalyzer, R

Indexation des données en utilisant **Elasticsearch**

Dashboarding par **Power BI**, Data visualization avec **nltk, ggplot**.

Environnement technique: **Python, R, JAVA, Spark, MySQL, Power BI, Spyder**

Libraries: **tweepy, pandas, numpy, nltk, os, findspark, SparkSession, pyspark.sql, searchtweets-v2, matplotlib.pyplot, plotline, ggplot**

Analyse Sémantique des Données Textuelles : recherche et innovation dans les technologies vertes de 2019 à aujourd'hui

[03/2023 – 04/2023]

Extraire les textes bruts des archives des actualités sur l'énergie à partir du site BFMTV en utilisant **BeautifulSoup, urllib.request, request**

Filtrage des textes par des mots pertinents de données textuelles avec **pandas, re** et **nltk stopwords**

Analyse, modélisation et interprétation des données textuelles, identifier des mots-clés et des tendances en utilisant **nltk.corpus**

Data visualization avec **seaborn, WordCloud, matplotlib**

Environnement technique: **Python, Jupyter Notebook**

Libraries: **urllib.request, request, pandas, re, nltk, BeautifulSoup, numpy, seaborn, WordCloud, matplotlib**

COMPÉTENCES NUMÉRIQUES

Database

MySQL / Neo4j / NoSQL / MongoDB / SQL

Langage de programmation

R / Python / C/C++

Big Data

PySpark / Elasticsearch / Kibana / Power BI / Spark / Base de données

Python

scikit-learn / NLTK / Natural Language Toolkit / pandas / numpy / multiprocessing / scipy / matplotlib / BeautifulSoup / request / seaborn / WordCloud

Machine Learning

Régression logistique / KNN / Réseaux de Neurones / Arbre de décision / Régression linéaire / Analyse en Composantes Principales (ACP) / SVM

ÉDUCATION ET FORMATION

Master 2 en Data Science, Mention: Innovation, Entreprise et Société

Université Paris-Nanterre [09/2021 – 09/2023]

Localité: Nanterre

Pays: France

Licence en Mathématiques et Informatique Appliquées aux Sciences Humaines et Sociales (MIASHS)

Université Paris-Nanterre [09/2018 – 06/2021]

Localité: Nanterre

Pays: France

LANGUES

Anglais TOEIC 895/990

Français B2 TCF 460

Vietnamien langue maternelle