

Anatole Lapellerie

Ingénieur en Data science

Mobile : 33 6 95 88 69 60 | Mail : anatole.lapellerie@gmail.com | LinkedIn : anatole-lapellerie | GitHub : Anatole-Lapellerie

Ingénieur Data passionné par les projets à impact, j'ai acquis une solide expérience en data engineering, data science et analyse de données dans des environnements à forts enjeux (banque, secteur public, e-commerce). Aujourd'hui, je souhaite mettre ces compétences au service de l'intérêt général, en accompagnant la transformation des politiques publiques par la donnée.

Compétences

Data science et IA en Python

- Data et visualisation (Pandas, Pyplot, Seaborn), Machine learning et Deep learning (Scikit-Learn, PyTorch)
- Computer vision (Torchvision, Yolo), Natural language processing (NLTK, GPT), algorithmes et Data mining (recherche en réseaux)

Big data et traitements

- Spark (PySpark, Scala), HDFS, extract-transform-load (ETL), automatisation et DevOps (pipelines, Shell, Jenkins, Git, Docker)
- Hive, SQL avancé (bases de données, requêtes complexes) et NoSQL (MongoDB, Neo4j, Redis)

Visualisation et analyse

- Dashboards (Data studio, Splunk, R shiny), analyses statistiques et reporting (Excel, analyses BI)
- Données géographiques et SIG (Mapbox, GeoJSON), développement web (HTML, CSS, PHP, JS)

Aptitudes : curieux, rigoureux, créatif, capacité de vulgarisation

Expériences

Consultant en Data — CGI — depuis 09/2023 — Paris

J'interviens en mission sur des sujets de Data pour des clients bancaires et financiers.

Data engineer — Société générale — depuis 02/2024 — Paris

Environnement : Spark, Scala, SQL, Hive, R, Jenkins

Au sein d'un projet international de nouveau TMS (transaction monitoring system) de conformité bancaire (lutte contre le blanchiment d'argent et le financement du terrorisme), je participe à la mise au point de plusieurs ETL (extract, transform and load) en Big data.

- Rétro-engineering des ETL déjà existants, réalisation de documentations et montée en compétence dans l'univers bancaire US et FR
- Développement de nouveaux ETL, de l'ingestion des fichiers transactionnels à la distribution des données dans des tables finales
- Réalisation ponctuelle de tâches de Data scientist et de Data mining au service d'un autre projet de conformité bancaire.

Data engineer — Caisse des Dépôts — 09/2023 à 12/2023 — Paris

Environnement : PySpark, Shell, SQL, Control-M, Jenkins

Au sein d'un ETL alimentant des vues métier, je suis chargé d'améliorer les traitements de données existants.

- Mise à jour de la documentation fonctionnelle des ETL
- Reporting quotidien de la Data quality et réalisation d'un job de cadrage pour simplifier les contrôles

Data scientist (stagiaire) — Société générale — 02/2023 à 08/2023 — Paris

Environnement : Python, R, Shiny, Spark, Scala

Pour un projet de lutte contre le financement du terrorisme, je participe au développement d'une application de Data mining en réseaux.

- Rétro-documentation de l'algorithme de Data mining, état de l'art de la théorie des graphes appliquée aux réseaux financiers criminels et aux réseaux sociaux, explication des choix techniques en place et création de livrables techniques
- Propositions et mise en place d'améliorations de l'algorithme de Data mining
- Maintenance de l'outil de data visualization et mise en place de nouvelles fonctionnalités graphiques

Data analyst (stagiaire) — Lacoste — 10/2021 à 03/2022 — Paris

Environnement : SQL, Google Cloud Platform (GCP), Data Studio, Excel

Au sein de l'équipe Data analytics & CRO de l'activité e-commerce mondiale, j'assiste les Data analysts.

- Mise à disposition de nouveaux dashboards Data Studio, Qlik Sense, Datama et Excel pour le suivi de KPIs
- Publication de rapports d'activités des sites à différents intervalles, à destination des responsables business
- Réalisation d'études pour expliquer les évolutions de performances sur des articles, catégories ou pays dans un contexte épidémique. Mise en relation des données Google Analytics et des données commerciales sur GCP

Data scientist (stagiaire) — ENGIE Digital — 06/2021 à 08/2021 — Paris

Environnement : Python, PyTorch, Yolo, Smart API, Amazon S3

Au sein du projet d'hypervision urbaine Livin', à destination des collectivités, je suis affecté à un projet de POC de Computer vision pour l'estimation du trafic urbain par caméra vidéo à l'aide d'intelligence artificielle.

- Définition du besoin et réalisation d'un état de l'art de la mesure innovante et non intrusive du trafic (théorie du trafic)
- Développement d'un POC de Computer vision grâce à un algorithme de tracking en direct (PyTorch, Yolo, Deepsort) et d'une API (Smart API) renvoyant les résultats sous forme de logs sur Amazon S3
- Développement d'un dashboard Splunk comprenant l'historique des résultats, la comparaison avec la théorie du trafic (diagramme de débit-densité), l'état du trafic en direct et la comparaison avec le streaming vidéo

Formations

École supérieure d'ingénieurs Léonard-de-Vinci (ESILV) — 2019 à 2023 — Paris

Diplôme d'ingénieur et Master (MSc) en Data science et en intelligence artificielle.

- Cours : Machine learning, Deep learning, NLP, Data mining, Visualisation
- Projets : Analyse de sentiments (TripAdvisor, NLTK), Chatbot NLP (JavaScript, Regex ; [voir GitHub](#))

University of Newcastle — 2022 — Newcastle, Australie

Échange universitaire d'un semestre, spécialisation en Data mining et en Deep learning.

- Projet : Classification d'images de produits industriels via CNN (PyTorch)

Lycée Saliège — 2017 à 2019 — Toulouse

Classes préparatoires aux grandes écoles (CPGE), spécialisation en physique et sciences de l'ingénieur (PSI).

- Projet : Sasport, un site web social pour organiser et suivre en direct les tournois sportifs du lycée

Personnel et loisirs

Français : langue maternelle

Anglais : **courant** (TOEFL ITP : 577 pts + semestre en Australie)

Allemand : **intermédiaire** (B2)

Sports : rugby à XV, ski alpin, randonnée (Pyrénées, NZ)

Loisirs : voyages (Australie, NZ, Océanie), lectures scientifiques (anthropologie...)

Projet personnel de site web social pour rassemblements amicaux. Développement full-stack, avec intégration de données time-series et GIS.