

imenlhocine@gmail.com

06 11 67 98 19

Formation

Master BigData et IA
- Université Paris8

Machine learning in Python with
scikit-learn
- INRIA

Sépcialisation en deep learning
- Coursera

Langues

Anglais: Professionnel

Imène

Data Engineer

Compétences clés

Techniques :

- **Langages** : Python, Shell, HTML/CSS
- **Manipulation des données** : SQL, Pandas, Spark, Scala
- **Visualisation des données** : Looker
- **Tests** : Black, Pytest, Unittest
- **API** : FastAPI
- **Prototype** : Flask, Dash
- **Cloud** : Google Cloud Platform, Amazon Web Services
- **Orchestration** : Airflow, Google composer, Google workflows
- **Gestion du code**: Git, GithubAction, gitlab
- **Gestion de l'environnement** : Docker, Terraform
- **Data warehouse** : BigQuery
- **Ingestion** : Talend, stitch, Airbyte

Fonctionnelles :

- **Méthodologie** : Agile/Scrum
- **Machine learning** : NLP, NLU, IA générative, prompt engineering
- **Communication** : Vulgarisation d'un sujet technique
- **Veille technologique** : Synthétiser l'état de l'art d'un sujet

Mes motivations

Passionnée par la data et l'ingénierie, je suis toujours à l'affût des nouvelles technologies concernant la data, l'IA et le cloud. Mon objectif est de construire des plateformes de données fiables et robustes en utilisant la meilleure architecture et des services managés du cloud, tout en étant attentive à la qualité et à la fiabilité des données.

Expériences professionnelles

DialOnce – Data engineer oct 2022 - sept 2024

Mise en place de la plateforme data de l'entreprise dans le cloud Google

Mise en place de la pipeline permettant l'ingestion, stockage et chargement de données (ELT) depuis des API REST jusqu'à BigQuery dans le cadre du lancement d'un nouveau produit

Développement et documentation de l'API de l'entreprise en utilisant les services managés de google.

LIASD Paris 8 – Data Scientiste avril-2022 - sept2022

Création d'un modèle qui catégorise les tweets haineux des non haineux à hauteur de 90% et déploiement du dashbord temps réel sur Heroku



DialOnce Data Engineer

Projet :

Créer une pipeline permettant l'ingestion, stockage et chargement de données (ELT) depuis des API REST jusqu'à BigQuery

Contexte :

DialOnce est un éditeur de logiciel, dans le cadre du lancement d'un nouveau produit cockpit omnicanal qui permet aux clients (BUT –TUI) de suivre et de piloter leur relation client à travers un dashboard en croisant les données clients de différentes sources .

Missions :

- Design, architecture, Mise en place de la pipeline de données
- Création de script de collecte des données depuis différentes APIs de nos clients
- Structuration des données dans GCS
- Mise en place des différents workflows d'orchestration des données avec google workflows , Scheduler et Composer
- Mise en place des différentes liaisons projets GCP, Dataform , github pour le versionning et l'isolation des données
- Conception et mise en place des différentes alertes à différents paliers pour assurer la continuité et qualité des données
- Automatisation de l'infrastructure as code via terraform
- Mise en place des environnement de production et de staging sur airflow

Environnement technique :

Python, Airflow Cloud Composer, GCS, Terraform, Bigfunctions, webhooks slack , procédure SQL , github



DialOnce Data Engineer

Projet :

Implémentation de la data plateforme sur le cloud Google

Contexte :

Au sein de DialOnce , éditeur de CRM spécialisé dans la digitalisation de la relation client et dans le cadre de la migration des données de l'entreprise vers le cloud GCP et l'adoption des différents services managés de google.

Missions :

- Ingestion des données de différentes sources
- Adoption de dataform pour la transformation des données
- Adoption de Looker et migration des visualisations de data studio vers Looker
- Automatisation des processus d'ingestion avec Google Workflows et Scheduler
- Mise en place de la chaine CI/CD entre Dataform et Looker
- Échange avec les équipes Google sur l'adoption des outils Looker et Dataform
- Formation et montée en compétences des équipes (data analyst) sur ces différents sujets

Environnement technique :

Google BigQuery, Airbyte, Talend (stitch), SQL, Google Workflow, Google Dataform, Google scheduler, YML, GithubAction, Lookml, looker, dbdiagram

DialOnce

Data Engineer

Projet :

Développement et documentation de l'API de l'entreprise en utilisant les services managés de google.

Contexte :

Le but de ce micro-service est de mettre à disposition de nos clients les KPIs les concernant via API et dépôt SFTP .

Missions :

- Mise en place des tests de sécurité de l'API pour s'assurer de la qualité et de l'isolation des données
- Documentation de l'API et sa mise à disposition sur le site de l'entreprise
- Conseil et onboarding des clients pour l'utilisation de l'API (ENEDIS - SOFINCO - GMF)
- Isoler la couche de données permettant d'exposer les Kpis dans Dataform puis dans Looker
- Développement d'un micro-service de dépôt des données récupérées via API sur un serveur SFTP

Environnement technique :

Python , Git , Bash, docker, Looker , Auth2, Notion , Rest Client

LIASD PARIS 8

Data Scientiste (6 mois)

Projet :

Création d'un modèle qui catégorise les tweets haineux des non haineux à hauteur de 90%

Contexte :

- Constitution d'un jeu de données français étiqueté automatiquement par des méthodes d'apprentissage non supervisé
- Entraînement d'un modèle de classification à partir de ce jeu de données

Missions :

- Extraction des données avec l'API Tweepy
- Nettoyage des données avec Pandas, Spacy, Nltk
- Analyse et visualisation avec Seaborn, Wordcloud, Matplotlib
- Vectorisation TfIdf , Word2vec, Transformers
- Apprentissage non supervisé (Clustering, Topic Modeling)
- Apprentissage supervisé (Modèle de classification et de régression)
- Détection d'anomalies (Isolation Forest)
- Validation de la solution accuracy, Matrice de confusion , cross validation

Environnement technique :

Python, NLP, Spacy, Pandas, Matplotlib, vectorisation , feature engineering , détection d'anomalie , clustering

Projet prsonnel

Data Scientiste

Projet :

Dashboard de détection de la haine sur Tweeter

Contexte :

Ce projet a été développé dans le cadre de la validation de la Licence informatique, il s'agirait de scraper plus d'un million de tweets , catégoriser ces derniers (grâce à une IA) et exposer le résultat sur un dashboard.

Missions :

- Connexion avec la base de données avec (SqlAlchemy, Flask)
- Base de données (PostgreSQL, PgAdmin4)
- Langage (Python , Css , Html5, Javascript, Pandas)
- Présentation des données avec les frameworks (Bootstrap, Plotly.js, Dash)
- Tests unitaires et test d'intégration (UNITTEST, PYTEST, SELENIUM)
- Déploiement du dashboard (HEROKU)

Environnement technique :

Python, Dash, SQL, pandas · Flask, Selenium, Pytest, Heroku, PostgreSQL

Projets divers

2019- 2024

Contexte :

- Implémentation de SOM en C from scratch
 - Gestion des bases de données : Sql, PostreSql, NoSql, PGAdmin4
 - Création d'un langage en SLY: lexer, parser, interpreteur (packagé :Pypi)
 - Création d'un shell en C: makefile, complétion, redirection, arrière_plan, pipe
 - Jeu de NIM vs IA en Python: en POO interface avec Pygame, IA
 - Jeu de Pierres Feuilles Ciseaux en Python en POO interface avec Pygame
 - Classification binaire et multiclasse
 - Mise en place d'architecture de données distribuées dans le cloud AMAZON (**EC2, VPC, S3, EMR, Cloud9, Hive, Hadoop, Pyspark, Jupyterhub**) permettant l'analyse de données (nework taxi trip) dans le cloud AWS
 - Prédire la note de touristes à partir de leurs avis
1. V 1: Python , analyse de sentiment , word2vec, TFIDF , NLTK, Vader, Blob, Pandas, matplotlib, classification multiclasse , classification binaire
 2. V2: Pyspark, MLlib , parquet

Environnement technique :

Python , C, Bash, pyspark , AWS