



Hamza EL BELGHITI

Étudiant en 2ème Année Master Sciences et Ingénierie des Données à la recherche d'un stage de projet de fin d'étude à partir du mois de février pour une période de 6 mois dans le domaine de l'ingénierie des données.

Hamagistral.me

Hamagistral

Hamza El Belghiti

hamza.lbelghiti@gmail.com

FORMATION

MASTER SCIENCES ET INGÉNIERIE DES DONNÉES | 2022-2024

Faculté des Sciences de Rabat (FSR), Rabat

Cours: Data Science avec Python, Business Intelligence, Datamining, Big Data, Deep Learning

LICENCE MATHÉMATIQUES APPLIQUÉES | 2019-2022

Faculté de Sciences et Techniques de Mohammedia (FSTM), Mohammedia

Cours: Bases de Données SQL, Analyse, Algèbre Linéaire, Probabilités et Statistiques avec R

EXPÉRIENCES PROFESSIONNELLES

DÉVELOPPEUR FULL STACK | Juillet - Septembre 2023

3D SMART FACTORY, Mohammedia | [GITHUB](#) [DEMO](#)

- Déploiement du modèle deep learning MeshSegNet qui effectue la segmentation des dents en 3D, à l'aide d'une API RESTful. L'API a été hébergée sur AWS Cloud en utilisant les services Lambda, ECR et API Gateway, ce qui a réduit le coût d'utilisation de 85 \$ par mois à 1,3 \$ pour 1000 requêtes.
- Création d'une interface utilisateur utilisant Vite, React.js, Three.js, et VTK.js, permettant aux utilisateurs de segmenter les scans intra-oraux 3D, télécharger, analyser et visualiser les résultats en 3D.
- Intégration de Firebase Auth pour l'authentification. Utilisation de l'API OpenAI pour créer un chat bot assistant et déploiement de l'application sur le web à l'aide de Vercel.

Technologies : Vite, React, Tailwindcss, Threejs, Vtkjs, Firebase, Pytorch, FastAPI, Docker, AWS ECR + Lambda + API Gateway, OpenAI, Vercel

RESPONSABLE DE TÂCHE FRONT END | Mai - Juin 2023

OMDNA ÎLE-DE-FRANCE, À Distance | [GITHUB](#) [DEMO](#)

- Collaboration avec une équipe de plus de 10 personnes dans le développement d'une application de chatbot qui aide à trouver des transports publics alternatifs pendant les grèves en Île-de-France.
- Utilisation de Streamlit et Folium Maps dans le frontend. Web scraping du site web RATP pour l'état du trafic en temps réel, et téléchargement des données vers le bucket AWS S3 avec un ETL.
- Développement d'un algorithme qui calcule la station de transport disponible la plus proche en fonction des préférences et de la localisation de l'utilisateur. Intégration de l'API de Navitia, garantissant une récupération à jour des données en communiquant directement avec les services RATP.

Technologies : Streamlit, Folium, BeautifulSoup, AWS S3 + Lambda, CloudWatch, Navitia API

PROJETS

ANALYSE DES DONNÉES D'UN COMMERCE DE DETAIL EN LIGNE | 2023

PROJET DATA ENGINEERING | [GITHUB](#) [DEMO](#)

- Conception et mise en œuvre d'une pipeline de données ETL à l'aide d'Airflow, optimisant les processus de données pour une société de vente au détail en ligne.
- Ingestion des données brutes dans un GCP Storage Bucket. Modélisation des données en schéma en étoile, transformation des données à l'aide de dbt et chargement des tables dans BigQuery.
- Mise en œuvre de contrôles automatisés de la qualité des données à l'aide de Soda, permettant d'améliorer la fiabilité des données et création d'un tableau de bord à l'aide de Looker Studio.

Technologies : Python, Astro, Airflow, dbt, Soda, GCP Storage, BigQuery, Docker, Looker Studio

ANALYSE D'OFFRES D'EMPLOI D'INGÉNIEURS DE DONNÉES | 2023

PROJET DATA SCIENCE & DATA ENGINEERING | [GITHUB](#) [DEMO](#)

- Web scraping des offres d'emploi d'ingénieurs de données depuis Glassdoor hebdomadairement puis stockage des données brutes dans un AWS S3 Bucket.
- Extraction des données depuis S3, nettoyage et transformation des données puis chargement des données nettoyées dans un autre S3 bucket grâce à une pipeline de données créée avec Mage.
- Analyse des données (EDA) et développement d'un modèle Random Forest Regressor de prédiction des salaires d'ingénieurs de données déployé sur le web via Streamlit Community Cloud.

Technologies : Python, Selenium, AWS S3, Boto3, Mage, Scikit-learn, Altair VizStreamlit

COMPÉTENCES

PROGRAMMATION

Python • Java • C • R

BASES DE DONNÉES

Oracle • SQLite • MySQL • PostgreSQL • MongoDB • Snowflake • Pinecone

DATA & MACHINE LEARNING

Pandas • Numpy • Seaborn • Matplotlib • Scikit-learn • TensorFlow • Keras • NLTK • Selenium • BeautifulSoup • Mage • dbt • Airflow • Spark • Kafka • Soda • PowerBI

DEVELOPEMENT WEB

HTML/CSS • Java/Typescript • MERN • Nextjs • Django • FastAPI • Streamlit

CLOUD

AWS • GCP • Firebase

OUTILS

Linux • Docker • Terraform • Git • Jira

LANGUES

Arabe • Anglais (C2) • Français (C2)

CERTIFICATS

DATA ENGINEER

Délivré par **DataCamp**

Septembre 2023

DATA SCIENTIST WITH PYTHON

Délivré par **DataCamp**

Décembre 2022

MACHINE LEARNING WITH PYTHON

Délivré par **FreeCodeCamp**

Avril 2022

CS50 WEB PROGRAMMING

Délivré par **HARVARD**

Février 2022

PARASCOLAIRE

STREAMLIT FOR EDUCATION

Étudiant Ambassadeur Leader | 2023-2024

INNOVASUS - FSR CLUB

Membre | 2023-2024

CENTRES D'INTÉRÊT

Blogging

Échecs

Hackathons

E-sports