



NADA BALKASSEM

Data Scientist Junior spécialisée en Machine Learning et modélisation prédictive, avec une expérience concrète en traitement de données massives et construction de pipelines end-to-end. Compétente en analyse de données, segmentation client et churn prediction. Expérience pratique en IA générative (GPT, RAG, prompt engineering) appliquée à des projets réels d'automatisation et d'analyse. Capacité à développer des solutions end-to-end en environnement Big Data, de l'analyse à la mise en production.



+212 627 059 222



nada.bel201@gmail.com



Nada Balkassem



Tanger, Maroc

Expériences professionnelles

Cloud Marketing Hub, Tanger

AI Engineer (Freelance)

Septembre, 2025 – Mai, 2026

- Développement et déploiement d'un pipeline automatisé basé sur l'API OpenAI GPT pour la classification de milliers de réponses emails (positives / négatives).
- Conception d'une stratégie de prompt engineering et mise en place d'un système d'évaluation via des métriques de Machine Learning.
- Extraction, scraping et normalisation de données issues d'emails marketing (liens, offres, métadonnées) afin de structurer les analyses.
- Mise en place d'un workflow de traitement batch optimisé intégrant caching et gestion du rate limiting pour réduire les coûts et améliorer les performances API.
- Implémentation de contrôles automatiques de validation et de cohérence des données garantissant fiabilité et reproductibilité des traitements.

Technologies : Python, Pandas, OpenAI API, SQL, Web Scraping

Data Scientist Stagiaire

Février, 2025 – Septembre, 2025

- Conception et développement d'un pipeline de Machine Learning end-to-end pour la prédiction de clients à forte valeur.
- Traitement et exploitation de plus de 400 Go de données via des outils d'ingestion et de requêtage analytique.
- Réalisation d'ingénierie de variables avancée et optimisation de modèles de Machine Learning.
- Utilisation de bibliothèques telles que Scikit-learn et Pandas pour l'entraînement et l'évaluation des modèles.
- Exploitation de bases de données analytiques (SQL, Apache Impala) pour la préparation et l'analyse des données.

Technologies : Python, Scikit-learn, Pandas, Apache Impala, SQL

PuertoTransit, Tanger

Stagiaire Full Stack Developer

Avril, 2023 – Août, 2023

- Développement d'une application de gestion client automatisant le cycle complet de traitement des dossiers, de la saisie à la facturation.
- Participation à la conception et au développement du back-end et du front-end de l'application.
- Mise en place d'interfaces utilisateur responsives et optimisées.
- Intégration de requêtes dynamiques et amélioration des performances applicatives.

Technologies : Laravel, PHP, SQL Server, Tailwind CSS, Ajax

Projets académiques

Chatbot intelligent de cybersécurité

- Développement d'un chatbot de cybersécurité répondant en temps réel aux questions sur les menaces via une architecture RAG, Gemini + base vectorielle FAISS.
- Génération de données synthétiques par modèles de diffusion pour enrichir la détection d'attaques réseau.
- Déploiement conteneurisé et scalable avec CI/CD Jenkins, Docker.
- Technologies: React, Flask, Docker, Kubernetes, MLflow, Diffusion Models, RAG, FAISS, Gemini

Détection de logiciels malveillants par Machine Learning

- Conception d'un système hybride combinant analyse de signatures et classification d'images conversion des binaires en représentations visuelles analysées par OpenCV.
- Entraînement et comparaison de modèles ML : Random Forest, SVM, Decision Tree évalués par F1-score et cross-validation.
- Technologies: OpenCV, Scikit-learn, FastAPI, Docker, Kubernetes

Prédiction du churn client (télécommunications)

- Développement d'une solution de streaming pour l'analyse en temps réel du comportement client.
- Mise en place d'un pipeline distribué pour l'ingestion et le traitement de données massives.
- Technologies : Apache Kafka, Spark Streaming, Flask, Cassandra

Formations académiques

Master en Sécurité IT & Big Data (2023 – 2025)

Faculté des Sciences et Techniques (FST), Tanger

Licence en Génie Informatique (2019 – 2023)

Faculté des Sciences et Techniques (FST), Tanger

Baccalauréat en Sciences Mathématiques (2018 – 2019)

Lycée Allal El Fassi, Tanger

Compétences techniques

Data Science & IA

- Construction de pipelines ML end-to-end (ingestion, feature engineering, modélisation, déploiement).
- Modélisation prédictive (classification, régression, clustering, segmentation client, churn prediction).
- Analyse statistique, EDA, data cleaning et analyse comportementale.
- Deep Learning (TensorFlow, PyTorch, Keras).
- Bibliothèques ML (Scikit-learn, Pandas, NumPy).
- LLMs et IA generative (GPT (OpenAI API), BERT, Hugging Face Transformers).
- Prompt engineering et évaluation de modèles de langage.
- Architecture RAG (Retrieval-Augmented Generation).
- Traitement et classification de texte.

Big Data & Data Engineering

- Traitement de données massives avec Spark, PySpark, Hadoop, Apache Impala.
- Streaming temps réel avec Apache Kafka et Spark Streaming .
- SQL avancé, bases de données (SQL Server, Cassandra, Impala).
- ETL et orchestration (Airflow).
- Web scraping et structuration de données non structurées.

MLOps & Déploiement

- Docker, Kubernetes pour déploiement et scalabilité.
- CI/CD, MLflow, DVC, monitoring de modèles.
- Optimisation des workflows et gestion des coûts API.

Développement Logiciel

- Langages : Python, SQL, Java, PHP.
- Backend : Flask, Laravel, FastAPI.
- Frontend: React.js, Angular, Tailwind CSS.
- Conception d'API REST et systèmes automatisés.

Data Visualisation & BI

- Power BI, Apache Superset.
- Matplotlib, Seaborn, Plotly.
- KPI, dashboards et reporting analytique.

Outils

- Git, Jupyter.

Qualités personnelles

- Organisation et gestion des priorités
- Autonomie et responsabilité
- Esprit d'analyse et sens du détail
- Esprit d'équipe et adaptabilité
- Communication efficace

Langues

- Arabe : Langue maternelle
- Français : Courant
- Anglais : Avancé

Centres d'intérêt

- Voyage
- Lecture