

YASMINA SAOUD

Data Scientist / Data Analyst

✉ saoudyasmina.ys@gmail.com ☎ +33 7 69 18 87 01 🔗 [LinkedIn](#) 📍 Béziers / Montpellier, France



PROFIL

Data Analyst / Data Scientist, Master MIASHS (mathématiques et informatique appliquées). Deux ans en environnement décisionnel sur un poste data polyvalent (analyse, intégration de sources multiples, modélisation) et cinq ans d'activité indépendante, où j'ai créé et développé ma propre activité. Je travaille les données textuelles et documentaires avec le NLP et les LLM (BERT, agents), et je conçois les pipelines qui les alimentent.

EXPÉRIENCE PROFESSIONNELLE

Data Analyst / Data Engineer, Groupe Angelotti 10/2024 - 10/2026
Alternance Béziers

- **Intégration de sources multiples**: développement de **connecteurs** d'ingestion depuis un ERP, des API/REST et des bases SQL ; réconciliation de référentiels hétérogènes (~1 440 tiers, ~300 opérations) avec **zéro perte de donnée critique** lors d'une migration ERP.
- **Analyses et recommandations pour la décision**: développement en Python d'un outil d'aide à la décision (ingestion multi-sources, pipeline de features, modèles de ****prédiction du risque de marge**** et de vitesse d'écoulement), restitué aux directions.
- **Pipelines ETL/ELT** sous SQL Server : conception, industrialisation, supervision quotidienne et optimisation des traitements (performance, montée en charge)
- **Analyse exploratoire et KPI** : une trentaine de rapports et tableaux de bord pour 5 services métier ; définition des indicateurs et des règles de calcul avec les équipes.
- **Documentation** des sources, traitements et transformations ; vulgarisation auprès d'utilisateurs non techniques.

Développeuse Web, Activité indépendante 2019 - 2024
Freelance Montpellier, France

- **Création et développement de ma propre activité** : prospection, relation client, chiffrage, livraison, en autonomie complète pendant 5 ans.
- Conception et mise en production de solutions web sur mesure pour des TPE et PME, du cadrage du besoin à la formation des utilisateurs

FORMATION

Master MIASHS : Mathématiques et Informatique Appliquées aux Sciences Humaines et Sociales, mention Bien

Université Paul Valéry Montpellier 3 · Montpellier, France · 2024-2026

Licence MIASHS, mention Assez Bien

Université Paul Valéry Montpellier 3 · Montpellier, France · 2020-2024

LANGUES

Français : Natif • Anglais : C1 • Arabe : C2 • Espagnol : B2

COMPÉTENCES

Python	SQL	ETL/ELT	Jupyter
Tableau	API/REST	Git	Scraping
PostgreSQL	Docker	R	SQL Server
MongoDB	LLM	ACP	Kafka
RAG	Streamlit	Agent IA	

CERTIFICATIONS

Generative AI with Diffusion Models • NVIDIA (2026)

PROJETS DATA & IA

- **Détection d'événements dans des articles de presse (NLP)**, (Python, BERT, INCEpTION)

Chaîne complète d'extraction d'information sur des textes journalistiques : définition d'un schéma d'annotation, annotation humaine, entraînement d'un modèle **BERT** en reconnaissance d'entités nommées (avancées technologiques, décisions publiques, conflits...), évaluation et application de visualisation des résultats.

- **Détection d'émotions par contexte visuel de scène** · projet de recherche de fin d'études, (PyTorch, YOLOv8, EfficientNet-B2, ResNet-50, Stable Diffusion)

Reproduction et extension d'un article de référence sur le jeu de données EMOTIC (23 571 images) : **ajout d'une branche de 114 features explicites et interprétables** (objets YOLO, couleurs dominantes, luminosité) à une architecture CNN double, avec fusion par attention et Focal Loss sur classes déséquilibrées, analyse critique des limites, et pipeline de **génération guidée** (Stable Diffusion) suivie de reclassification.

- **Challenge Deep Learning MIASHS 2026 1re Place**
Architecture **deep learning multimodale** (imagerie satellite + données climatiques) conçue et entraînée sous forte contrainte de temps. Conception d'un **agent LLM** répondant aux questions techniques du projet et génération automatique de rapports d'aide à la décision.

- **Pipeline de données en temps réel (qualité de l'air)** (Python, Redpanda/Kafka, Apache Flink, PostgreSQL, Docker, Streamlit) : Architecture Kappa complète : connecteur d'ingestion (API Open-Meteo), traitement en flux avec fenêtres temporelles et gestion du temps d'événement, écriture en **PostgreSQL**, tableau de bord temps réel. Infrastructure reproductible en une commande (``docker compose``), évaluation de la latence.