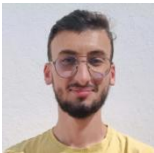


BEN TOUHAMI Mohamed Rida

Data Engineer

AL-Hoceima, Morocco |+212 650-421-062 bentouhamimohamedrida@gmail.com



PROFILE SUMMARY

Data Engineer skilled in designing scalable data pipelines and delivering actionable insights through analytics, with a passion for solving complex problems using Data Science and emerging technologies.

EDUCATION

- National School of Applied Sciences of Al Hoceima, Abdelmalek Essaâdi University 2020 - 2025
Diploma in Data Engineering | Relevant Coursework: Databases, Big Data, Machine Learning, Cloud Computing.
- Moulay Ali Chérif High School, Baccalaureate in Mathematical Sciences A 2019 - 2020

WORK EXPERIENCE

DIGITAL PLACE COMPANY

Data Science Intern

Tangier, Morocco
Feb 2025 – Jun 2025

- Built a computer vision pipeline for semi-trailer identification using YOLOv11, PaddleOCR, and Grounded SAM2 for license plate detection, OCR, and segmentation.
- Trained YOLOv11 models for license plate detection, achieving high precision (0.9779, 0.990) and recall (0.9625, 0.987).
- Fine-tuned PaddleOCR for Arabic and European license plates with 93.75% and 87.5% accuracy
- Built and fine-tuned an **EfficientNet-B2 classification model** to distinguish Moroccan (Arabic script) and European (Latin script) license plates, achieving 100% validation accuracy with Adam optimizer.
- Utilized **DINOv2** to generate numerical vector representations (vector embeddings) of semi-trailer components (truck, container, cab)
- Developed a FastAPI and React app, containerized with Docker, and deployed via a **Bitbucket CI/CD pipeline** on a Hostinger VPS for automated integration and deployment.

Tech stack: Python, PyTorch, FastAPI, React, Docker, Bitbucket, YOLOv11, PaddleOCR, Grounded SAM2, DINOv2, EfficientNet, Git

SHIFTBRICKS COMPANY

Data Engineering Intern

Al Hoceima, Morocco
Jun 2024 – Sep 2024

- Designed a data pipeline using Medallion architecture to process Arabic legal documents into JSON schemas, storing metadata and embeddings in MongoDB for AI applications.
- Applied OCR (Tesseract, easyOCR) and LLM models (e.g., Cohere) for extraction.
- Automated workflows with Airflow, and built a FARM stack UI (FastAPI, React) for expert validation.
- Containerized the solution using Docker for scalability.

Tech stack: Python, MongoDB, Airflow, FastAPI, React, Docker, Git, GitLab, OCR, GenAI.

TECHNICAL SKILLS

- Languages:** Python, Java, SQL, PL/SQL, Spring Boot
- Big Data & BI Tools:** Spark, Hadoop, Kafka, Airflow, Power BI
- Databases:** MySQL, Oracle, PostgreSQL, MongoDB, Cassandra, ClickHouse
- DevOps:** Git, GitLab, GitHub, Docker, Kubernetes
- Machine Learning:** Classification, Clustering, Regression, scikit-learn
- Cloud Services:** Databricks, Cosmos DB, Azure Synapse Analytics, Azure Data Lake, Azure Data Factory, Azure Functions

LATEST PROJECTS

AI Trends Chatbot on Azure with RAG and Medallion Data Pipeline Architecture

- Built a RAG chatbot with an automated Medallion data pipeline on Azure.
- Scraped AI trends using Azure Functions and BeautifulSoup, stored in Cosmos DB (bronze, silver)
- Processed embeddings in Databricks, indexed in AI Search (gold)
- Automated pipeline with Data Factory, deployed FastAPI (Cohere API) on Container Apps, and hosted Next.js frontend on Static Web App

Tech Stack: Azure Functions, AzureCosmos DB, Azure Databricks, AI Search, Data Factory, Container Apps, FastAPI, Cohere API, Next.js, TypeScript, BeautifulSoup, Medallion, RAG

Real-Time Apartment Data Processing Pipeline

- Ingested and streamed apartment data with Kafka, stored raw text in HDFS (Parquet)

- Extracted structured info using Cohere API, processed data in real time with PySpark Streaming
 - Stored transformed data in ClickHouse for fast SQL queries, with real-time visualizations via React.js & FastAPI
- Tech Stack:** Hadoop HDFS, Apache Kafka, PySpark Streaming, Cohere API, ClickHouse, React.js, FastAPI

Real Estate Price Prediction & Recommendation System (2nd Place – ML Competition, Oujda, Morocco)

- Scraped real estate data using BeautifulSoup and Selenium
 - Built a hybrid regression model and recommendation system with Python (Scikit-learn, pandas)
 - Developed a Flask web app with HTML/CSS & JavaScript to display predictions and suggestions
- Tech Stack:** BeautifulSoup, Selenium, Python, Scikit-learn, pandas, Flask, HTML/CSS, JavaScript

Agricultural Patent Analysis Data Pipeline

- Collected patent data from Google Patents and other sources and stored in MongoDB.
 - Transformed data with PySpark, loaded into PostgreSQL data warehouse, and analyzed with Power BI
 - Developed a Flask-based UI for data interaction.
- Tech Stack:** MongoDB, PySpark, PostgreSQL, SQLAlchemy, Power BI, Flask

CERTIFICATIONS

- Big Data with Spark and Hadoop Essentials (IBM),
- Data Warehousing and BI Analytics Essentials (IBM)
- ETL and Data Pipelines with Shell, Airflow, and Kafka (IBM),
- NoSQL Databases Essentials (IBM)
- Machine Learning with Apache Spark (IBM)

SOFT SKILLS

Proven teamwork in collaborative projects, adaptability, problem-solving, communication

LANGUAGES

French (Advanced), English (Advanced), Arabic & Tamazight (Native)