BEN TOUHAMI MOHAMED RIDA

DATA ENGINEERING STUDENT

About Me :

Enthusiastic data and software engineering student at the National School of Applied Sciences of Al Hoceima, driven by a passion for IT, data, and data science. Eager to learn and contribute to challenging projects, I possess strong foundational skills in data engineering/science and software development. Committed to continuous growth, I am currently seeking an end-of-year internship to apply my knowledge and skills in real-world scenarios and further develop my expertise.

CONTACT

+212650421062

bentouhamimohamedrida@gmail.com

AL HOCEIMA,MOROCCO

Open to remote work

@BenTouhamiMR.github

@BenTouhamiMR.portfolio

@BenTouhamiMR.linkedIn

TECHNICAL SKILLS

- **Big Data & BI Tools**
  - Spark - hadoop - Data Warehouse - kafka - Airflow - Web Scraping(BeatifulSoup/ Selenium) - Power BI
- **Machine Learning**
  - classification - clustering - Regression - scikit-learn
- **Deep Learning**
  - NLP - CNN - RNN -LSTM-Keras
- **DataBases**
  - Oracle - Mysql - Cassandra - MongoDB- SQLServer - PostgreSQL
- **Programming languages & Framworks**
  - Python - Spring boot(java) - Spring Data JPA - React - C - PL/SQL Html - Css - javascript- Shell
- **Operating System**
  - Linux (Ubuntu) - Windows
- **Version Control**
  - git - github

LANGUAGES

- **Arabe** (Native)
- **French** (Advanced)
- **English** (Intermediate)

CERTIFICATES

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

EDUCATION

- **Data Engineering & Preparatory cycle**
  - 2020-2025 National School of Applied Sciences of Al Hoceima
- **Baccalaureate of Science in Mathematics**
  - 2019-2020 Molay Ali Chrif High School, Al Hoceima

PROFESSIONAL EXPERIENCES

AI Data Ingestion Pipeline for AI Applications - Internship at Shitbricks Startup

- **Pytesseract (OCR):** Used for extracting data from PDFs.
- **Cohere API:** Extracts metadata from the content of files, structures the files, and applies embeddings to chunks of structured text.
- **MongoDB Atlas (cloud):** Stores the structured files, metadata, and embeddings.
- **FastAPI:** Backend of the application for human validation.
- **React:** Frontend of the application for human validation.
- **Airflow:** Automates the pipeline from the bronze layer to the gold layer.
- **Postgres:** Stores validation logs.
- **Docker:** Containerizes the application.

ACADEMIC AND PERSONAL PROJECTS

Patent Analysis Project

- **BeautifulSoup & APIs** : Used for collecting patent data from web pages and various APIs.
- **MongoDB Atlas (Cloud)** : For storing the collected patent data.
- **Apache Spark** : For data analysis and processing of the stored patent data.
- **Postgres(Data Warehouse)** : For storing transformed patent data in a structured format for further analysis.
- **SQLAlchemy** : For managing database operations and interactions with PostgreSQL.
- **Power BI** : For visualizing insights and analysis from the data.
- **Flask Web Application** : For building the user interface where users can search for patents and select specific patents for further analysis.

Real Estate Apartment Price Prediction and Recommendation Project (2nd Place in ML Competition - Oujda, Morocco)

- **BeautifulSoup/Selenium:** Used for web scraping to collect real estate data from web
- **Python (Scikit Learn, pandas, etc.):** For data preprocessing and building machine learning models.
- **Hybrid Regression Model:** Predicts property prices based on the collected data.
- **Recommendation System:** Suggests similar apartments based on user preferences.
- **Flask & HTML/CSS & JavaScript:** building a user-friendly web interface to display
- predictions and recommendations.

Design and Development of an Exam Scheduling Application

- **Spring Boot:** Used to build the backend of the application.
- **Spring Data JPA:** For managing and interacting with the database.
- **React:** For building the frontend, enabling user interaction and scheduling management.
- **MySQL:** For storing exam schedules and related data
- **Docker**: containerizing the application and simplifying deployment across environments