Sami RAJICHI

Data Scientist

Mahdia, Tunisia

📞 +216 27944801   ✉ semi.rajichi@gmail.com   in linkedin.com/in/sami-rajichi/   ⊙ github.com/sami-rajichi

About me

Passionate, collaborative, dedicated professional with strong project delivery and leadership for innovation. Eager to contribute a blend of skills and unwavering commitment to excellence in the realm of data science.

Professional Experiences

Wind Consulting 🌐   **Dec 2024 – Present**

Data Scientist - Mahdia, Tunisia - Remote

- **Fine-tuned LLAMA 3.1** on the company's custom dataset, **converted it to GGUF**, and integrated it into **a containerized API** for seamless deployment.
- Developed an **ERP RAG chatbot** as an assistance guide and integrated it into a containerized API.
- Trained a **deep learning model** using **multilingual embeddings model** on AR, EN, and FR dataset for advanced NLP task, and integrated it into a containerized Django API powering a **rule-based chatbot** for a travel agency.
- Engineered an end-to-end document scanner: **localized CIN fields** via OpenCV/JSON, then deployed a containerized Django API **integrating EasyOCR** for automated **text extraction** with enhanced Arabic handling and side validation logic.
- **Technologies/Tools: Python, LLM, Transformers, Tensorflow, PyTorch, RAG, OLLAMA, LangChain, API, Django, Docker, EasyOCR**

SamiX, Master's Graduation Project 🛆   **Feb 2024 – Nov 2024**

Intern in Data Science - Mahdia, Tunisia - Remote

- Developed a **virtual voice assistant** with **6 services** (programming, reports, search, general tasks, I/O control, presentation) for **hands-free automation**.
- Fine-tuned the **Whisper ASR model** on **5,000+ self-recorded audios** to match **personal speech patterns**.
- Implemented a **personalized eye-tracking** solution, enabling **gaze-based interaction** for users with limited mobility.
- **Technologies/Tools: Python, HuggingFace, Tensorflow, Kaggle, Gradio, Tkinter, NLP, LLM, Computer Vision**

Shamash-IT 🌐   **Jul 2023 – Sep 2023**

Intern in Data Science - Tunis, Tunisia - Remote

- Performed in-depth **Exploratory Data Analysis**, addressed **outliers** to ensure data accuracy, and created **impactful visualizations** using Power BI to reveal key patterns and trends. ⊙
- **Technologies/Tools: Python, Power BI, Github**

Capabl 🌐   **Aug 2021 – Sep 2021**

Intern in Machine Learning - Bengaluru, India - Remote

- Developed an **inventory management system** ⊙ and honed **data analysis** skills through **Summer Olympics dataset** ⊙, while demonstrating commitment to healthcare research by **building a neural network for early cancer detection**. ⊙
- **Technologies/Tools: Python, CNN, ML Algorithms.**

Freelance   **Dec 2023 – Present**

Data Scientist - Remote

- **Dual-Stage Graph Neural Network (DSTAGNN) for Traffic Forecasting**: Implemented a **two-stage GCN-LSTM** architecture to capture both spatial and temporal dependencies in urban traffic data, achieving highly accurate speed predictions.
- **Class Activation Maps (CAM) for Skin Cancer Analysis**: Implemented **Grad-CAM** to generate and visualize **activation maps**, highlighting critical regions in skin cancer images.
- **Skin Cancer Classification using Vision Transformer**: Designed and implemented a **ViT-based model** to classify skin lesions.
- **Data Augmentation for ASAYAR Dataset**: Applied **diffusion models** with Image-to-Image and Text-to-Image approaches to expand the **ASAYAR dataset**.
- **DICOM Image Segmentation**: Developed a **custom CNN model with transfer learning** for DICOM image segmentation, delivering precise, high-quality outputs for clinical and research use.
- **Technologies/Tools: Python, Kaggle/Colab, Keras, ViT**

Education

Protected Consulting (NACS-Certified) ⊕   Oct. 2024 – Dec. 2024
Master of Business Administration (MBA-TIC)   *Tunis, Tunisia*

Higher Institute of Computer Science of Mahdia ⊕   Sep. 2022 – Nov. 2024
Professional Master in Data Sciences (MP-SD)   *Mahdia, Tunisia*
- **Honors**: Admitted with High Distinction; Valedictorian of the Class.

Higher Institute of Computer Science of Mahdia ⊕   Sep. 2019 – Jun. 2022
Bachelor of Computer Science, Specialization in Software and Information Systems Engineering   *Mahdia, Tunisia*
- **Honors**: Admitted with High Distinction; Valedictorian of the Class.

Projects

Web Knowledge Assistant ⊙
- Engineered a cross-platform Python application leveraging **Crawl4AI** and dual scraping techniques (**Markdown** & **LLM via Groq API**) for efficient web crawling and advanced content processing.
- Integrated a dynamic **Gradio** interface featuring **real-time** content previews and an interactive chat system powered by **LangChain**, **FAISS**, and **ChatGroq**.

Live Audio Capture – Python Package ⊕
- Developed a cross-platform Python package using **FFmpeg**, **noisereduce**, and **PyQtGraph** for **real-time** audio capture, **noise reduction**, and **voice activity detection**.
- Integrated advanced features like **adaptive thresholding**, **low-pass filtering**, and **real-time visualization**.
- Published on **PyPI** with automated **CI/CD** pipelines via **GitHub Actions** and comprehensive documentation using **MkDocs**.

ChatSMA ⊙
- Developed a genetics and SMA disease **chatbot** using **Generative AI (Llama 2)** for personalized conversation.
- Integrated **Langchain for natural language processing** and **Pinecone for cloud-vector database management**, and designed user-friendly interface using **Flask**.

Face Attribute Detection ⊙
- Utilized cutting-edge **deep learning techniques** and **transfer learning models** such as Xception, EfficientNetB7,...
- Achieved exceptional results with over **95% accuracy in gender prediction**, a remarkable **99.7% accuracy in face mask detection**, and an impressive **89.8% accuracy in glasses or no glasses classification**.

HumanForYou: Art of Human Sustainability ⊙
- Leveraged **analytical techniques** to identify and address key drivers of **employee turnover** at HumanForYou, achieving a **99% precision with RandomForest** by **balancing dataset** using **SMOTE**.
- Optimized model accuracy using **model turning techniques** like GridSearchCV and RandomizedSearchCV.

ScLakhas: Health and Genetics Text Summarizer ⊙
- Engineered a **text summarization system** catered to genetics scientific texts, using the **LAKHAS architecture**.
- Applied **text normalization techniques using NLTK** to improve data preprocessing and model accuracy.

Skills & Competencies

- Skilled in machine learning and AI, including NLP, computer vision, large-scale data processing, and generative AI.
- Strong communicator with leadership abilities, adaptable and collaborative in dynamic environments.

Languages

- **Arabic**: Native
- **English**: Professional
- **French**: Intermediate

Achievements / Extracurricular Experiences

PEEMo - Student Entrepreneur Center Monastir   Apr 2023 – May 2023
Candidate
- Achieved the "Initiator" status

Google Developer Student Clubs - Mahdia   Aug 2021 – Aug 2022
President
- Led my team to achieve **1st rank in Tunisia and 3rd rank in the Middle East North Africa region**.

AIESEC Ruspina   Feb 2021 – Oct 2021
Business-to-Customer Team Member
- Enhanced **communication abilities** through active engagement with diverse audiences

ATAST ISIMa club   Oct 2020 – Nov 2021
President
- Enhanced key **soft skills** during this experience, notably improving **communication and teamwork abilities**