HOUMAN RAJABI

Data Scientist

Contact

+39 351 9465199

houmanrajabi@myyahoo.com

houmanrajabi.github.io

in/houmanrajabi

github.com/houmanrajabi

huggingface.co/houmanrajabi

Via Borselino, 42, 10138, Turin, Italy

Technical Skills

Languages & Core: Python (OOP), SQL, Bash.

AI & GenAI: LLMs, RAG Pipelines, LangChain, Hugging Face, Vector Search, Fine-tuning (PEFT/LoRA).

Machine Learning: PyTorch, TensorFlow, Scikit-learn, XGBoost, Computer Vision, NLP.

MLOps & Cloud: Docker, Kubernetes, AWS (SageMaker), CI/CD (GitHub Actions), MLflow, Airflow.

Big Data: PySpark, Kafka, Databricks, Data Warehousing.

About Me

Data Scientist & Machine Learning Engineer focused on deciphering complex systems to deliver actionable insights. I bridge the gap between raw data and strategic decision-making by combining the rigor of statistical modeling with the engineering of LLMs and RAG systems. Nearing graduation from the University of Turin, I offer a perspective sharpened by professional collaborations across Italy, Germany, and Iran.

- **Tech Stack:** Python, SQL, Statistical Modeling, and Adaptive Machine Learning (from GLMs to Transformers).
- **Global Experience:** Successfully collaborated with diverse teams in Italy, Germany, and Iran.
- **Core Focus:** Delivering end-to-end actionable insights.

EXPERIENCE

Data Scientist Intern | Trivago
Düsseldorf, Germany | 08/2025 – 01/2026

Data Science Intern in the central team, architecting the End-to-End ETL Pipelines for Unstructured Data Ingestion.
- Leveraged Generative AI (NLG) to automate the creation of accommodation descriptions and Computer Vision to optimize image quality.
- Built semantic search engines using embedding models to improve query-to-accommodation matching accuracy.
- Engaged in research-driven development, ensuring methodologies met rigorous scientific standards.

Data Scientist | Snapp! Tehran, Iran
March 2021 – July 2023

- Forecasting: Developed models to balance driver supply and demand, reducing wait times by 2 minutes.
- Algorithms: Designed dynamic pricing algorithms to maximize revenue (+3%) during high-demand periods.
- Machine Learning: utilized real-time data to enhance ETA prediction accuracy by 2.5%.
- Analytics: Monitored KPIs to identify and execute optimizations that improved driver utilization and completion rates.

Projects

open-market-intelligence
(RAG & Multimodal AI)
Developed a multimodal RAG pipeline for financial analysis using Qwen2-VL for vision-based table extraction and vLLM for semantic chunking, incorporating sandboxed code execution to render accurate plots instead of hallucinating visuals.

role-sync
(Agentic AI & Orchestration)
Built a human-in-the-loop resume analysis agent using LangGraph and Flask to map skills and identify gaps interactively.

evalita-multipride-2026
(NLP Research & Fine-tuning)
Designed a hybrid NLP framework (BERT + MLP) for detecting slur reclamation across three languages, utilizing language-specific feature fusion.

LANGUAGES

English
- Full professional proficiency

German
- Limited working proficiency

Italian
- Elementary proficiency

CERTIFICATIONS
- AWS Certified Machine Learning – Specialty

Data Scientist | Digikala Group Tehran, Iran
May 2019 – March 2021

- Engineered hybrid recommendation engines (Association Rules, Collaborative Filtering) to redesign the "Frequently Bought Together" module, boosting Average Order Value (AOV) by 5% via cross-selling.

- Developed price optimization algorithms for Flash Sales ("Shegeftane") to automate candidate selection and discount depth, achieving ~95% sell-through without eroding margins.
- Built robust demand forecasting pipelines (XGBoost, Prophet) capable of handling 4x traffic loads during Black Friday, reducing logistics bottlenecks by 10% and improving inventory accuracy.
- Enhanced search relevance by integrating custom Persian NLP models for semantic matching and misspelling correction, reducing "Zero Search Result" queries by 7%.

Collaborations

Researcher & Data Scientist | University of Turin (DeepHealth Project)
Sep 2023 - May 2025
- **Federated Learning:** Architected decentralized training workflows across hospital networks to avoid moving raw patient data.
- **Technical Liaison:** Aligned clinical requirements with HPC infrastructure for industrial partners like Philips and Thales.
- **Privacy Engineering:** Deployed GDPR-compliant algorithms to ensure data sovereignty within strict hospital IT environments.

Project Contributor | Sharif University of Technology
Sep/2018 – Mar/2019

- Institutional Analytics: Created robust analytical dashboards to support institutional decision-making and enhance strategic academic planning.

EDUCATION

Master of Science in Language Technology & Digital Humanities (NLP)
University of Turin | Sep 2023 - Jun 2026

Bachelor of Science in Computer Science
Sharif University of Technology | Sep 2015 – Mar 2019

PUBLICATION

Identity, Toxicity, or Complexity? (EVALITA 2026)

1st Place (Italian Task): Developed a Hybrid Fusion architecture combining BERT embeddings with engineered sociolinguistic features to detect reclaimed slurs; achieved SOTA F1 of 0.8981 by modeling language-specific syntactic patterns.

Contact

+39 351 9465199

houmanrajabi@myyahoo.com

in/houmanrajabi

github.com/houmanrajabi

huggingface.co/houmanrajabi

Via Borselino, 42, 10138, Turin, Italy