

ISMAIL AHOUARI

Email: ismailahouari123@gmail.com Mobile: +393455310181 Location: Milan, Italy
LinkedIn: [linkedin.com/ismailahouari](https://www.linkedin.com/ismailahouari) GitHub: github.com/SameulAH

SUMMARY

Data Scientist and Deep Learning enthusiast pursuing a Master's in Data Science at the University of Milano-Bicocca, with a strong foundation in applied mathematics and machine learning. Dedicated to advancing expertise in emerging technologies, with a particular interest in the intersection of data science and infrastructure to develop scalable, real-world AI solutions.

EXPERIENCE

Luxembourg Institute of Socio-Economic - LISER Jan 2025 - June 2025
Data Science Intern

- Implemented a scalable multilingual semantic classification pipeline using pandas, and Polars enabling efficient processing of large-scale text data for NLP tasks.
- Used Data preprocessing modules for HTML content extraction(BeautifulSoup), text normalization, and deduplication with spaCy improving input quality for downstream machine learning models.
- Integrated Stanza for language-specific sentence segmentation and developed modular components to support multilingual NLP processing across diverse corpora.
- Developed a keyword extraction system using Sentence-Transformers (Hugging Face) with semantic similarity techniques to identify AI-related indicators within domain-specific corpora.
- Optimized pipeline throughput using batch processing, data parallelism, and language-specific branches, reducing latency and improving model inference speed.
- Benchmarked the semantic similarity pipeline against GPT-based models (OpenAI GPT, Mixtral) to assess performance accuracy and approach fidelity, contributing to model selection for real-world AI applications.

University of Luxembourg – C2DH June 2024 - Dec 2024
Data Science Intern

- Deployed a Wikibase-based knowledge graph using Docker containers for isolated, reproducible environments
- Automated data ingestion pipelines (Python & Wikibase API) to process 10K+ structured records with semantic annotations
- Developed a modular relational data model managing RDF triples across 19 reusable properties
- Created multilingual support system with automated translation capabilities for Luxembourgish accessibility
- Established data reconciliation workflows using OpenRefine for cross-platform entity linking
- Built SPARQL query interface enabling complex historical research queries and data visualization

Numidia Collection Nov 2021 - Feb 2022
Web Dev Intern

- Developed and managed CMS platforms, including WordPress and Prestashop.
- Built and optimized websites using HTML, CSS, and JavaScript, ensuring compliance with W3C standards.
- Handled backend development for specific projects using Django and Flask.
- Implemented SEO strategies to improve search engine rankings and enhance website visibility.

EDUCATION

Master Degree in Data Science Aug 2022 – Sep 2025
University of Milan- Bicocca, Milan, Italy

Diploma in Web Technologies Feb 2020 – Dec 2021
EWA School, Agadir, Morocco

Bachelor Degree in Applied Mathematics Aug 2016–Nov 2020
University of Ibn Zohr, Agadir, Morocco

SKILLS

Languages: Python , C / C++ , SQL , R

Data Tools: MLflow Apache Spark, Docker, Airflow, SQL Server

Libraries & Frameworks: PyTorch, TensorFlow, Scikit-learn, Pandas, Polars

Other: Tableau, Git, Flask, AWS

PROJECTS

• Data Management & Visualization

- Conducted a data management project analyzing Spotify music trends by collecting track metadata via the Spotify API and scraping song lyrics from external sources.
- Performed data cleaning, matching, and preprocessing to align song attributes with lyrical content.
- Applied TF-IDF to extract key lyrical themes and uncovered patterns related to artist popularity and evolving music trends.
- Built a Neo4j graph database to model relationships between songs, artists, and themes, enabling network-based exploration of musical trends.

• Marketing Analytics

- Analyzed customer behavior data to inform data-driven marketing strategies and retention efforts.
- Performed RFM segmentation to cluster customers based on value and engagement.
- Built predictive models using Random Forest, Logistic Regression, and Decision Trees to forecast repurchase likelihood and churn.
- Applied the market basket analysis to uncover product affinity patterns.
- Implemented review sentiment classification for customer feedback analysis.

• Technological Infrastructure for Data Science

- Implemented a multi-layered data lake architecture on Azure Data Lake Storage using bronze, silver, and gold zones to support scalable and structured data refinement for analytics.
- Developed automated ETL pipelines in Azure Data Factory to orchestrate seamless data ingestion from on-premises SQL Server to the Azure cloud data platform, enabling centralized data access.
- Built scalable data transformation workflows using Azure Databricks and PySpark, enabling distributed processing and aggregation of datasets for downstream consumption.
- Created dynamic, interactive dashboards in Power BI with real-time data connectivity via DirectQuery, delivering actionable insights and KPIs to business stakeholders.
- Established a monitoring and alerting framework using Azure Monitor, Log Analytics, and Data Factory Alerts, ensuring pipeline reliability, proactive failure handling, and data quality assurance.

• Digital Signal & Image Processing

- **Music Classification:** Analyzed audio signals dataset (MFCCs, spectrograms) and applied SVMs and neural networks to classify music tracks and genres.
- **Food101 Classification:** Classified images across 101 food categories using CNNs and transfer learning for model optimization.
- **Image Retrieval:** Developed an image retrieval system by extracting features with CNNs and implementing similarity-based(KD-Tree) search algorithm.
-

CERTIFICATIONS

1. AWS Certified Solutions Architect - Associate – <i>in progress</i>	<i>August 2025</i>
2. MLops Zoomcamp – <i>in progress</i>	<i>August 2025</i>
3. IBM Professional Data Scientist	<i>Sep 2022</i>
4. IBM Professional Data Analyst	<i>March 2022</i>

LANGUAGES

Native: Berber, Arabic

Other language(s): English , French, Italian