Leon Doungala

AI / Machine Learning Engineer - Data Scientist

Milan, Italy • (+39) 32 93 67 00 86 • doungala.leon@gmail.com • linkedin.com/in/leon-doungala • Portfolio : Link [AI & ML / GenAI and RAG]

AI/ML Engineer - Data Scientist with an M.Sc. in Computer Engineering (AI specialization) and a Master's in Data Science, dedicated to delivering innovative solutions. Proficient in predictive and generative AI, with expertise in LLMs, RAG, Transfer Learning, and Vector Databases. Preparing for registration with the Italian Order of Engineers (Ordine degli Ingegneri), Section A, in 2025.

EXPERIENCE

AI / ML and Data Science Engineer *March 2024 - Present*

People Lab - *Milan, Italy*

- Developed AI solutions for 300+ clinics and companies, including a RAG-based Clinical Semantic Search Engine and an ATS system for HR recruiters supporting 200+ companies, using LangChain, **Flask, Scikit-learn, PyTorch**, and OpenAI LLM API.
- Designed AI projects with Generative AI, RAG, Transfer Learning, LLMs, and Machine Learning, building predictive systems and scalable solutions using **Deep Learning**, MongoDB, MySQL, Azure Cosmos DB, **Python**, and Vector Databases.
- **Built AI systems** for **semantic search, automation,** and **autonomous agents** in healthcare and industry.
- Led AI projects from concept to deployment, ensuring efficiency, scalability, and impact.

Software Developer | Data Scientist *April 2022 – January 2024*

Italian Stock Exchange (Euronext Group), *Milan, Italy*

- Developed and maintained financial software applications, optimized **SQL DataMarts** (**Workbench, Athena**), and automated AWS data ingestion and processing with **Python**.
- Collaborated within **agile teams**, managed AWS data ingestion (Glue, Python Spark), and enhanced machine learning models through **data mining, preprocessing, and model building** for **risk assessment**.
- Applied **data science techniques** to support **quantitative analysis, decision-making,** and **financial forecasting**.

Developer Analyst Salesforce *Jun 2020 – December 2021*
Lutech group , *Milan, Italy*

HCM Technical Consultan Intern *Sep 2019 – Feb 2020*

Oracle , *Milan, Italy*

EDUCATION

eCampus University - *Come, Italy* *April 2025*
Master of Science (M.Sc.) in Computer Engineering and Automation

Rome Business School, *Rome, Italy* - **University of Valencia**, *Valencia, Spain* *April 2022*
Double Specialized Masters in Data Science

Ferrara University - *Ferrara, Italy* *September 2020*
Bachelor's Degree in Electronic and Computer Engineering

SKILLS

- **Programming and Scripting**: Python, R, SQL, C++, Java
- **Machine Learning**: Random Forests, Gradient Boosting, SVM, Decision Trees, K-NN, Logistic Regression, XGBoost , Ensemble Learning etc …
- **Deep Learning Frameworks**: TensorFlow, PyTorch, Keras
- **Generative AI** : Langchain , Vector Databases , Embedding , Semantic Search , RAG (Retrieval - Augmented - Generation) , LLMs models (OpenAI , Anthropic)
- **Cloud Platforms**: AWS, Google Cloud, Azure
- **Software development & Engineering** : Flak , Docker , Java EE

CERTIFICATIONS

Ferrara Innovation Consortium *2019 - 2020*
- *Advanced Techniques for Design and Development of Industry 4.0 Web Applications and Digital Transformation (Duration: 1 year)*

Michigan University *November 2022*
- *Certificate in Python Programming*

ENSAE - ENSAI formation continue *December 2023*
- *Certificate of dashboarding with tableau*

LANGUAGES : French (C2 - Native) , Italian (C1- Proficient) , English (B2)

M.Sc. THESIS AND PERSONAL PROJECTS

Personal Portfolio and GitHub links for more details : https://leondoungala22.github.io/doungala.leon.github.io/ml_and_ds_portfolio.html

Project Name	Focus	Key Points	Repo
M.Sc. thesis : Artificial Intelligence and Machine Learning Techniques for the Characterization and Prediction of Diabetes	Applied AI and ML techniques, including RAG, semantic search, and generative AI, to address diabetes prediction and characterization challenges. Developed supervised (e.g., **XGBoost**, **Random Forest**) and unsupervised (e.g., **K-Means**, **PCA**) models for actionable medical insights. **Langchain** , **Keras**. Developed a full Advanced AI web app for Diabetes Monitoring	• Designed integrative dashboards and predictive analytics for healthcare. • Developed advanced predictive models for diabetes risk and characterization. • Implemented **LangChain** and Vector Databases for RAG workflows and semantic search. • Leveraged Generative AI for data augmentation and handling imbalanced datasets.	**Private** [Access on request] https://github.com/LeonDoungala22/Msc-Thesis-AI-ML-for-Diabetes-Characterization
Virtual Healthcare AI Assistant	• Description: The AI Health Assistant is an AI-powered system designed to assist individuals by providing helpful, reassuring, and "medically" sound guidance based on their symptoms . **Using Langchain , Flask , OpenAI API models , Anthropic** • Key Features: Dynamic medical data retrieval, conversational flows, and task automation.		Public [Free Access] https://github.com/LeonDoungala22/AI-Health-Assistant
ATS System for CV Filtering and Job Matching	• Description: Developed an AI-powered ATS system to align candidates with job descriptions or assist HR with advanced semantic search. Utilized **RAG**, **LangChain**, **Anthropic APIs**, **OpenAI**, and **Azure Cosmos DB** for precision matching and intelligent filtering. • Algorithm: Retrieval-Augmented Generation (RAG) with semantic search and conversational AI.		Public [Free Access] https://github.com/LeonDoungala22/RAG-ATS-Semantic-matching-search-cv-ai-v1.0.0
Human Resources Analytics	• Description: Applied PCA and K-Means clustering for employee segmentation and HR data analytics. • Algorithm: **PCA, K-Means**		
Sentiment Analysis on IMDb Movie Reviews	• Description: Performed sentiment analysis to classify reviews using Naive Bayes techniques. • Algorithm: **Bernoulli Naive Bayes**		Public [Free Access] https://github.com/LeonDoungala22/Practical_AI_ML_DL_DataScience_Portfolio/blob/main/code/Module%202_Scikit-Learn%2C%20Basic%20ML%20Models%2C%20EDA%2C%20and%20ML%20Practice%20Projects/Practical%20projects%20(%20ML%20models%20)%20%20Naive%20Bayes/Sentiment%20Analysis%20on%20IMDb%20Movie%20Reviews/Bernoulli%20Naive%20Bayes/main.ipynb
Breast Cancer Classification	• Description: Developed a classification model for breast cancer using K-Nearest Neighbors. • Algorithm: **KNN**		Public [Free Access] https://github.com/LeonDoungala22/Practical_AI_ML_DL_DataScience_Portfolio/blob/main/code/Module%202_Scikit-Learn%2C%20Basic%20ML%20Models%2C%20EDA%2C%20and%20ML%20Practice%20Projects/Practical%20projects%20(%20ML%20models%20)%20%20K-Nearest%20Neighbors%20(KNN)/Breast%20Cancer%20Classification%20with%20K-Nearest%20Neighbors%20(KNN)/main.ipynb
Heart Disease Prediction	• Description: Built predictive models to analyze heart disease risks using ML and NLP techniques. • Algorithm: **SVM, NLP**		Private [On review] https://github.com/LeonDoungala22/Practical_AI_ML_DL_DataScience_Portfolio/blob/main/code/Module%202_Scikit-Learn%2C%20Basic%20ML%20Models%2C%20EDA%2C%20and%20ML%20Practice%20Projects/Practical%20projects%20(%20ML%20models%20)%20%20SVM%20(%20Linear%20and%20non%20linear%20)/Heart%20Disease%20Prediction/main.ipynb
Credit Card Fraud Detection	Credit Card Fraud Detection" project aims to develop a decision tree classifier to detect fraudulent credit card transactions. The project involves importing and exploring the credit card fraud dataset, handling missing values, and balancing the class distribution between fraudulent and non-fraudulent transactions. Key technologies used in this project include: • Python libraries: **Pandas, Matplotlib, Seaborn,** and NumPy for data manipulation and visualization. • **Scikit-Learn** for machine learning model development and evaluation. • Missingno for visualizing missing data. • Custom functions and additional libraries like Handler and pickle .		Public [Free Access] https://github.com/LeonDoungala22/Practical_AI_ML_DL_DataScience_Portfolio/blob/main/code/Module%202_Scikit-Learn%2C%20Basic%20ML%20Models%2C%20EDA%2C%20and%20ML%20Practice%20Projects/Practical%20projects%20(%20ML%20models%20)%20%20/Decision%20Trees/Credit%20Card%20Fraud%20Detection/main.ipynb