

# MIGUEL FERREIRA DA SILVA

Lisbon | [miguel.marinho.silva@outlook.com](mailto:miguel.marinho.silva@outlook.com) | +351 933 852 719 | [LinkedIn](#) | [mmarinhofs.vercel.app](https://mmarinhofs.vercel.app)

## WORK EXPERIENCE

### NOKIA, Lisbon – Portugal

#### Data & Analytics - Data Scientist Lisbon – Portugal

Sept 2022 – Current

- In charge of creating & enhancing complex supply chain data business objects (**Spark SQL**) that can be used by business to facilitate end-to-end reporting, transforming their business requests to data engineering solutions
- Designing and implementing critical **supply chain KPIs** (milestones) for operational reporting
- Utilizing **PySpark RDDs** and **DataFrames** for automation & data manipulations in **Databricks** environment.
- Developed an advanced text generation Large Language Model (**LLM**), integrating metadata from data objects for facilitated information retrieval. Coded in **Databricks** workspace, using **MLflow**, **transformers**. **ETL** with **Spark**.
- Scaled file comparator to **compare tables** across multiple servers, streamlined **UAT** with precise mismatch diagnostics & integrated with **Azure DevOps** for Continuous Integration and Continuous Deployment (**CI/CD**)
- **Azure Cloud Machine Learning studio**: Predicting On-Time Delivery of Supply Chain Products, stacked ML model
- Migrated data pipelines from on-premises DB to **Azure SQL DB**; implemented solutions using **Azure Data Factory**
- Automated extraction, processing, and ingestion of Celonis audit logs **API** data into data pools using **Python**. Developed star schema, associated tables (**SQL**), & dashboard.

## EDUCATION

### Católica Lisbon School of Business and Economics, Lisbon – Portugal

Sept 2021 – July 2023

*Masters in Business Analytics – Specialization in **Data Science*** - GPA: 17/20

- **Python**, DB Management (**SQL**), Big Data Technologies (**Hadoop**, **Spark**, **Hive**)

**Thesis: Deep Learning for Melanoma Classification: A Study Using Skin Lesion Images** - 19/20

- MAR-MELA-CNN Fusion Ensemble trained on MAR-MELA dataset containing 8000 Melanoma images.
- Introduced the **F $\beta$  Score** for melanoma detection, minimizing false negatives in critical diagnoses.
- Scaled model for real-time inference, delivering instant melanoma diagnosis via SMS from mobile photos.

### Nova School of Business and Economics, Lisbon – Portugal

Sept 2017 - June 2021

*Bachelors Economics & Data Analytics*

### European School of Luxembourg I

Sept 2005 - June 2017

*European Baccalaureate* - Multilingual education in Portuguese and English, French as complementary language.

## SKILLS

- **Languages:** Portuguese (C2), English (C2), French (C1), Spanish (B1), German (A1- *Learning*)
- **Programming Languages:** Python, PySpark, R, SQL
- **Machine Learning / Deep Learning:** scikit-learn, TensorFlow, Keras, Natural Language Processing: transformers, huggingface, datasets, MLflow
- **Data Analysis:** Pandas, NumPy, SciPy, Matplotlib, SeaBorn
- **Data Engineering:** Git, Databricks, Azure Data Lake, Azure SQL DB, Azure Data Factory, ETL/ELT pipelines
- **Platforms:** Azure, Azure Machine learning Studio, Azure DevOps, SSMS, Oracle, Databricks, SAP S4/Hana
- **Certifications:** [AZ-900 : Azure Fundamentals Certificate](#) / [DP-100: Azure Data Scientist Associate](#) / [Celonis \(8\)](#) / DeepLearning.ai: [Generative AI with Large Language Models](#), [Finetuning Large Language Models](#),