MIGUEL FERREIRA DA SILVA

Lisbon | miguel.marinho.silva@outlook.com | +351 933 852 719 | LinkedIn

WORK EXPERIENCE

NOKIA, Lisbon – Portugal

Data & Analytics - Data Scientist Lisbon – Portugal Sept 2023 – Current

Data Semantics Team

• **ECHO Δ:** Scaled File Comparator to **compare tables** across multiple servers, featuring a user interface for customizable **SQL** queries using **tkinter**. To enhance scalability and performance, integrated **Dask** for dataframe operations, optimizing both speed and handling of large datasets.
Significantly **streamlined the User Acceptance Testing (UAT) process** by providing precise mismatch diagnostics in cloud migration.
Integrated with **Azure DevOps** for continuous integration and continuous deployment (**CI/CD**)
Presentation to Data & Analytics department.

Data & Analytics - Data Science Intern in Discovery Program, Lisbon – Portugal Sept 2022 – Aug 2023

Supply Chain Advanced Analytics Team

• **ML** classification project using **Azure Cloud Machine Learning studio**: Predict delivery date in supply chain product
High understanding of Azure-specific packages & tools for model saving, assessment, and deployment.
Accompanied setup of Machine learning workbench, **AutoML** to automatically **gridsearch**, feature importance analysis and selection
• Migrated data pipelines from on-premises DB to **Azure SQL DB**; implemented solutions using **Azure Data Factory**

Process Mining Team

• **Data Engineering project-** Assisted in building a data pipeline from **Celonis API** to **Azure Function Apps**, automating daily data extraction and API updates
• **Dashboard Automation-** Designed and implemented automated scripts and pipelines for extracting, processing, and integrating **Celonis audit logs API** data into data pools using **Python**. Developed star schema, associated tables (**SQL**), and dashboard of Celonis Adoption in Nokia.
• **Machine Learning Project**- Descriptive analysis using **Matplotlib**, overfitting alleviation, **feature selection**, **hypertuning** to find optimal parameters using extensive grid search, mitigate class imbalance through **ADASYN/SMOTE** using **scikit-learn**, choosing appropriate metric, stacked models to achieve maximum F1 score. Developed final report dictating results and conclusions.

Data Semantics Team

• **SQL Validation-** Contributed on data migration tasks from on-premises server to Azure cloud, doing User Acceptance Testing (UAT) with **SQL** to ensure no information was lost in the migration
• **Excel File Comparator**- Lead in developing a **Python**-based **Excel File Comparator** program, ensuring data consistency between on-premise servers and Azure cloud by precisely locating and reporting any discrepancies. Presentation to Data & Analytics department.

Financial Planning and Reporting Analytics

• **Data Object Creation**: Lead in optimizing SAP report generation by recreating complex reports using advanced **SQL** queries on a centralized data platform with ingested SAP tables, significantly reducing processing time. Developed **Python** scripts for automated validation of reports.

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 **(**Pandas, Matplotlib, NumPy, SeaBorn, Git), Data Base Management (**SQL**), Big Data Technologies (**Hadoop, Spark, Hive**)
- Machine / Deep Learning: Executed a customer churn classification project using an ensemble of SVM/KNN/RF algorithms via **Scikit-learn**; Conducted an image classification project employing Convolutional Neural Networks (CNNs) with **TensorFlow**.

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

- Set up & deployed multiple **Azure** servers for parallel training of state-of-the-art MAR-MELA-CNN Fusion Ensemble of custom merged MAR-MELA dataset containing 15000 Melanoma images.
- Introduced **Fβ Score** metric for medical applications to balance the trade-off between precision and recall, prioritizing the minimization of false negatives for critical diagnoses.
- **Scaled model for real-time inference**, integrated with Google Drive, to apply prediction on image uploaded from my phone and **deliver instant melanoma diagnosis via SMS**

Nova School of Business and Economics, Lisbon – Portugal Sept 2017 - June 2021

Bachelors in Economics

SKILLS

- **Languages:** Portuguese (C2), English (C2), French (C1), Spanish (B1), German (A1)
- **IT:** Python – Skicit-Learn,Tensorflow, Dask, R, SQL, Tableau, Azure, Azure DevOps, Celonis, Hadoop, Spark, Hive, SSMS, Oracle

Certifications: AZ-900 : Azure Fundamentals Certificate / DP-100: Azure Data Scientist Associate Certificate / Celonis (8)

ACTIVITIES AND INTERESTS

- **Cultural multi diversity:** I communicate daily with various people from European nationalities. I have always been **open-minded** in getting to know other realities and cultures. I love travelling, highlighting my trip to Asia, where I fell in love with Eastern culture.
- **Technology:** Highly connected to this sector from a young age. **Self-taught** assembly of a computer from scratch on my own, with over 40 hours of research to obtain highest performance at smallest market price.
- **Boxing:** Develops my **resilience**, my **determination**, **discipline** and **ability to stay calm under pressure**, due to having to meet with the various restrictions that occur during times of competition. Goal to be national champion in this category. **10+ hours** of weekly training,
- **Chess:** Develops **strategic** thinking, analytical skills, decision-making abilities, **prepositioning** against possible adverse scenarios.
- **Kaggle ML:** Competing in various data science Machine Learning competitions honed skills in **data analysis**, **predictive modeling**