MAYKON JONATTAN DE OLIVEIRA

Junior Data Scientist | Data Analyst

Natal, RN, Brazil | maykinvest@outlook.com

linkedin.com/in/maykon-oliveira-maykinvest | github.com/MaykInvest | maykinvest.github.io/maykon-portfolio

SUMMARY

Data professional transitioning from 10 years in IT into data science, with freelance experience analyzing client datasets in Python, building machine learning models (Scikit-learn, XGBoost, LightGBM), and delivering Power BI dashboards to non-technical stakeholders. IBM Data Science Professional Certificate holder, completing a Postgraduate Certificate in Data Science. Fast learner; native Portuguese, proficient English.

SKILLS

Programming & Tools: Python, SQL, R, Git, GitHub, Jupyter Notebook

Data Analysis: Pandas, NumPy, Data Cleaning, Exploratory Data Analysis (EDA), Feature Engineering

Machine Learning: Scikit-learn, XGBoost, LightGBM, Classification, Regression, Clustering (K-Means), Hyperparameter Tuning, SHAP

Big Data & AI: Apache PySpark, LLM APIs (Groq), Streamlit

Data Visualization: Power BI, Matplotlib, Seaborn, Folium, Dashboards

EXPERIENCE

Freelance Data Scientist & Web Developer **2018 – Present**

Upwork | *Remote*

- Cleaned, processed, and analyzed complex client datasets in Python (Pandas, NumPy), uncovering business trends that informed clients' strategic decisions.
- Built and evaluated predictive models, including classification algorithms, with Scikit-learn to solve client-specific business problems.
- Created visual reports and dashboards that communicated findings clearly to non-technical stakeholders.
- Managed projects end to end, from requirements to deployment, delivering on time and within budget.

IT Help Desk Technician **08/2016 – Present**

Home for Poor Children (philanthropic institution) | *Mossoró, Brazil*

- Write SQL queries to extract data from internal databases and build interactive Power BI dashboards that track support metrics and reveal root causes of recurring issues.
- Provide hardware troubleshooting, software maintenance, and user support.

PROJECTS

Titanic Survival Prediction **Python, Scikit-learn, SHAP**

- Built a leakage-free classification pipeline with Bayesian hyperparameter tuning; the Gradient Boosting model reached 83% accuracy on unseen data (baseline 61.7%), explained with SHAP.

House Price Prediction **Python, LightGBM, XGBoost**

- Built an end-to-end regression pipeline on ~21,000 home sales comparing Ridge, Random Forest, XGBoost, and LightGBM; the best model cut cross-validated MAE by ~51% versus the baseline.

Energy Consumption Forecasting **Apache PySpark**

- Built a distributed PySpark pipeline forecasting hourly electricity demand from time-series and weather data; a Random Forest outperformed the Linear Regression baseline.

Car Market Segmentation **Python, K-Means**

- Segmented ~11,000 car models into three market groups with K-Means, choosing the number of clusters by silhouette score.

AI Python Programming Assistant **Python, Streamlit, LLM API**

- Built an AI chat app on the Groq LLM API that answers Python questions with explanations and code examples.

EDUCATION

Postgraduate Certificate in Data Science – Unopar University 03/2026 – Expected 10/2026

Postgraduate Certificate in Cloud Computing – Unopar University 02/2024 – 11/2024

Associate Degree in Cybersecurity – Unopar University (Online) 10/2020 – 06/2023

CERTIFICATIONS & LANGUAGES

IBM Data Science Professional Certificate (Coursera) | **Statistics & Probability** (Khan Academy)

Languages: Portuguese (Native), English (Proficient)