Rafaela Souza Pinter

+1 438 528-3449 | +55 48 99931-9933
rafaelaspinter@gmail.com
LinkedIn | GitHub

EDUCATION

Master of Science in Computer Science   09/2023 – 05/2025
Université de Montréal, Montréal – Canada

Specialization in Data Science & Big Data   02/2021 – 07/2022
Universidade Federal do Paraná (UFPR) – Brazil   *Final grade: 9.9/10*

International Exchange in Génie Biologique   09/2018 – 06/2019
Aix-Marseille Université (AMU), Marseille – France

Bachelor Degree in Bioprocess Engineering and Biotechnology   02/2015 – 08/2020
Universidade Federal do Paraná (UFPR) – Brazil   *Final grade: 8.7/10*

SKILLS

Languages:
- Native Portuguese, Fluent English, Fluent French, Basic German

Code Versioning:
- Git, GitLab, GitHub

CI/CD:
- GitHub Actions, GitLab CI/CD

REST APIs frameworks:
- FastAPI, Flask

Containers:
- Docker

Cloud:
- **Google Cloud Platform** (Vertex AI, Cloud Run, Cloud Build, Container Registry, Cloud Storage, BigQuery, App Engine)

Programming languages, frameworks, and tools:
- **Python** (Pandas, TensorFlow, Keras, Scikit-learn, Numpy, Kedro, XGBoost, CatBoost, Pyomo, Seaborn, Plotly, Plotly Dash, Streamlit, NLTK, requests, SQLite, DuckDB, Beautiful Soup, BentoML, FastAPI, Flask, Pytest, MLFlow)
- **SQL** (BigQuery, PostgreSQL, MySQL)
- **Java**
- **R**
- **C/C++**
- **HTML/CSS**

PROFESSIONAL EXPERIENCES

Machine Learning Engineer (11/2022 – 08/2023)
Falconi – São Paulo, SP – Brazil
- Developed a Machine Learning pipeline template with **MLOps** principles incorporating a **Kedro pipeline**, **Docker**, **MLFlow** tracking and a **CI/CD pipeline** for seamless **deployment on GCP's Vertex AI**, resulting in reduction in productization time for machine learning projects from months to days.
- Development of **unit and integration tests** with the **Pytest** framework to ensure quality of a Data Science modeling pipelines.
- Created a highly accurate CatBoost model for predicting delivery fees, achieving an annual MAPE of 14% and a trimester MAPE of 3%.
- Conducted data analysis of fuel resale prices for a major distributor in Brazil, identifying potential opportunities worth 40 Mi Reais.
- Constructed **CI/CD pipeline to automate the deployment** of **MLflow server**, **Streamlit**, and **Plotly Dash applications** on **GCP's App Engine**.
- Defined and implemented DevOps best practices for a team of 12 data scientists, facilitating efficient collaboration and integration of development and operations.
- Provided **guidance and expertise to a team of six data scientists in cloud development**, leveraging various tools and services such as Cloud Storage, BigQuery, Vertex AI Workbench, Vertex AI Pipelines, App Engine, and Cloud Run.

Data Scientist (08/2021 – 11/2022)
DP6 – São Paulo, SP – Brazil
- Implemented an end-to-end (data preprocessing, modeling and reporting) marketing mix model **(Robyn, LightweightMMM) pipeline in Python** to provide comprehensive insights into marketing effectiveness and optimize resource allocation.

- Implemented a **continuous training pipeline of a BigQuery** ML Time-series Forecasting (ARIMA) model to accurately predict transaction volumes during the Black Friday week, enhancing sales forecasting capabilities showing results in live Data Studio dashboard. The model predicted the Black Friday sales with an overall **97% accuracy**.
- Designed and implemented an **ETL** flow between **Smartsheet and Google Cloud Platform**, leveraging Cloud Functions to store HR data in BigQuery. With the database created, conducted analysis on over 1600 recruitment processes using **NLP** techniques such as TF-IDF and word2vec. Identified common characteristics among successfully hired candidates, enabling more informed HR evaluations.
- Enhanced a Next Best Offer **XGBoost** recommender model served to over 3 million clients using Gitlab and **GCP's AIPlatform**. Incorporated MLOps and CI/CD pipelines to the project to ensure seamless testing, model deployment, and updates.
- Implemented a budget optimization algorithm in Python, utilizing nonlinear optimization techniques (**Pyomo, Scipy**) to reallocate marketing budget effectively and optimize ROI.

Data Analyst (12/2020 – 07/2021)
Radar Saúde – São José, SC – Brazil
- Used an **NLP** algorithm to **automate** the **identification of duplicated** records within the Data Warehouse tables, significantly improving data accuracy and integrity.
- **Developed in AWS** an interactive **dashboard** using **R Shiny App** for women's health, offering comprehensive insights into pregnancy monitoring, breast and cervical cancer screenings, and HPV vaccine administration. This dashboard served as a valuable tool for healthcare professionals and stakeholders.
- Enhanced the functionality and usability **of Power BI dashboards** utilized by the Brazilian Public Health System (SUS), enabling more effective data visualization and analysis.

Levteck Tecnologia Viva – Florianópolis, SC – Brazil
Student Intern, Laboratory Technician (03/2020 – 11/2020)
- Implemented Plackett-Burman Experiment Design in bioreactors to conduct experiments, resulting in a statistical analysis that led to a significant 50% increase in company productivity specifically for yeast strains.
- Collected, cleaned and analyzed production data.
- Designed and structured an **SQLite database**, leveraging **Python for data analysis**, to investigate the impact of fermentative parameters on cell yield and bioreactor contamination.
- Conducted a comprehensive study on contamination patterns using **Python**, successfully identifying the source of bioreactor contamination through thorough analysis and investigation.

VOLUNTEER WORK

Mulheres em Dados (*Women in Data*), Brazil
Volunteer Mentor (08/2022 – 12/2022)
- Shared experiences and mentored women in the process of career transition to Data Science.

Colégio Estadual Padre Claudio Morelli, Curitiba-PR – Brazil
Volunteer Teacher (08/2017 – 11/2017)
- Created the extension project "Future Freshmen". Teacher of tutoring in mathematics, physics, and chemistry for senior students of a public high school.

CERTIFICATIONS

- Machine Learning modeling Pipelines in Production (DeepLearning.AI, 2023)
- Machine Learning Data Lifecycle in Production (DeepLearning.AI, 2023)
- Introduction to Machine Learning in Production (DeepLearning.AI, 2023)
- MLflow: management of ML models life cycle (Alura, 2023)
- MLOps: Machine Learning and APIs (Alura, 2022)
- Data Warehouse Concepts, Design, and Data Integration (University of Colorado, 2020)
- CS50's Introduction to Computer Science (HarvardX, 2020)