PEDRO MORAES - MACHINE LEARNING & AI ENGINEER
e-mail: pedromoraes20@gmail.com   linkedin: www.linkedin.com/in/pedrovmoraes/

I am Pedro Moraes, Brazilian, and a resident of Porto-PT. I have a decade of experience in **Big Data**, **Data Science**, and **Machine Learning** initiatives, leading the development of scalable data-driven systems for finance, logistics and customer service companies and also delivering algorithms in clouds **Azure** and **AWS** and taking care of the life cycle management of these models. Currently working with **Generative AI,** and **LLM models** and researching **alternative data** applied in investments.

PROFESSIONAL EXPERIENCES

CLOSER CONSULTING | Senior Data Scientist *(Jan/2023 - Jan/2024)*

Consultancy project at Rubix:

Rubix is a reseller of factory products. With thousands of professionals doing business in 32 countries, the company chose to invest in artificial intelligence tools to help salespeople be more productive, so I worked with LLMs, OpenAI GPT 3.5-turbo and GPT4 models, and Azure Machine Learning to achieve that.

Main tasks:

➢ I was responsible for the MPV for using LLMs to read and analyze thousands of emails daily. I designed the architecture of the pipeline in Python packages, set up the infrastructure for Azure ML workspace and its environment for research development, and deployed it to production.

➢ I set up GPT 3.5-turbo and GPT4 to be used in the project. I implemented the pipeline in Azure Machine Learning, and also applied NLP libraries and OpenAI libraries such as tiktoken to better deal with the amount of tokens.

➢ I created and managed Docker files for keeping the libraries to read, process, and save data before and after applying LLMs.

➢ I developed, tested, and monitored the creation of prompts to be applied to the GPT model calls.

➢ I made different deployments on Azure Machine Learning, in combination with Promp Flow, a tool to better create AI applications and deployments as APIs.

➢ I did technical management of a small team: 2 Data Analysts, 1 Data Engineer.

Tech stack: OpenAI, LLMs, NLP, Azure Machine Learning, Databricks, Docker, Prompt Engineering, and Azure DevOps.

Consultancy project at Real Madrid C. F.:

Real Madrid has an Applied Science team that develops several projects, including data from players, fans, and the business itself. I helped this team to predict visitors to the club's Stadium and Museum, this facilitated the planning of the marketing and sales teams.

Main tasks:

> ➢ I set up the Databricks workspace, and Delta live tables looking at databases with data from marketing campaigns and fans.

> ➢ I coded the Forecasting model for Stadium and Museum in Azure Databricks, using SparkSQL to query data, Python to create features and model and set the pipeline to run weekly.

> ➢ The artifacts from the models were saved in MLFlow and the code in Azure Devops.

> ➢ I used the Databricks dashboard to show the metrics of the models and results for predictions.

Tech stack: Databricks, MLFLow, Delta Live Tables, PySpark, Azure DevOps.

RADIX ENGINEERING AND SOFTWARE | Machine Learning Specialist *(Jul/2021 - Dec/2022)*

Consultancy project at Raizen (Cosan & Shell):

Raizen is a joint venture between the British company Shell and the South American company Cosan. They have hundreds of gas stations and revenue of around $20B. They need to predict the demand for gas stations, to not stop delivering gas anywhere. They also wanted to predict the cashflow, in such a way as to allow them to make short investments with that. I worked on these two projects.

Main tasks:

> ➢ I worked on migrating data pipelines for train and making predictions from Airflow to Azure Machine Learning.

> ➢ I improved the pipelines by creating Python packages, modularizing the tasks, hiding sensitive information, setting Docker image and Azure ML environment, and deploying taking advantage of Azure ML pipelines.

> ➢ Improved forecasting models to consider Neural Prophet as a new option of a model.

> ➢ Automated the deployment of new models with MLFlow, moving new versions of the model to production by automatically analyzing metrics of the new models against the previous one.

Tech stack: Airflow, Azure Machine Learning, Python, Docker, Github.

Consultancy project at Electrolux:

Electrolux is a global producer of kitchen products. And it has an e-commerce where it sells its products directly to end consumers. To improve marketing campaigns, I used e-commerce and customer data in a Customer Lifetime Value algorithm.

Main tasks:
> ➢ I designed the entire implementation of the CLV model: model selection, feature engineering, MLFlow management, and Azure function deployment.
>
> ➢ I was responsible for sharing results, also presenting development in Reviews, and explaining results and model definitions to C-Level people.
>
> ➢ I managed the team: 1 Data Scientist Jr, and 1 Data Engineer.

Tech stack: Azure Databricks, Azure functions, Python, MLFlow, CLV.

AMBEV TECH | Data Scientist *(Jan/2021 - Jun/2021)*

Ambev Tech is a technology subsidiary of AB Inbev, the largest beer company in the world. With millions of restaurants as customers, in the financial area, the Brazilian subsidiary Ambev acts as a credit agent for its partners, providing products with different terms. Each of the millions of restaurants has a credit score. I worked on this algorithm.

Main tasks:
> ➢ I worked on the entire development of the Credit Score model, selecting models, defining data architecture, and coding the model itself.
>
> ➢ To create features I need to process millions of payments all the time, so Databricks was the step to pre-processing data and creating features.
>
> ➢ The model was developed in different steps, by adapting the code in Databricks notebooks by running: model training and model selection, saving artifacts in MLFlow as OOP development. And then deploying everything as a pipeline in Databricks.

> ➢ I was also responsible for sharing results, presenting development in Reviews, and explaining results and model definition to C-Level executives.

Tech stack: Databricks, MLFlow, Credit score, PySpark, Python.

HEXAGON AGRICULTURE | Data Scientist *(Jul/2018 - Dec/2020)*

Hexagon is a Swedish software company, which does business in different areas. The Agriculture division has software and hardware for managing and assembling large fleets, as well as integrated management solutions for the Supply Chain. I worked on researching and developing data science analysis and machine Learning models (forecasting and clustering) for IoT (Internet of Things), using terabytes of sensor data coming from tractors and trucks (temperature, velocity, engines).

Main tasks:
> ➢ I first worked with the product manager on design MVP for cloud machine learning consuming: AWS Sagemaker, pricing, cost optimization, and deployment definition.
>
> ➢ I worked on the consumption and pre-processing of big data for applying ML models with PySpark and AWS Lambda.
>
> ➢ Then I worked on the first models using velocity from IoT devices on the trucks to make forecasting about delivery time. I also tested many other models using data like temperature and pressure.
>
> ➢ I worked on the deployment of these models in combination with the main product of the company - Fleet management.
>
> ➢ In reviews, I used to present the development and define the roadmap for new big data analyses and machine learning models.

Tech stack: AWS, AWS Lambda, Amazon Sagemaker, Amazon Quicksight, Git.

PLONUS | Co-founder *(Jul/2015 a Jun/2018)*

Plonus was a startup created by me and two friends. Started the business focused on image analytics for buildings like roads, dams, and bridges. Nowadays the company does, in addition to geodesic projects, environmental studies.

Main tasks:
> ➢ Product definition and development.
>
> ➢ Sharing results with partners and customers.

Education:

CEU - Central European University - Master visiting student in Economics and Business department (2022);

Universidade de São Paulo (USP): MSc in Economic USP University - Quantitative Analysis of Risk in Capital Marketsm(2023);

MIT on Edx - MicroMaster in Data Science & Statistics (2021).

Agricultural & Mechanical Engineering - Bachelor (2013).

Languages:

Português - Native

English - C1

Spanish - B1