Rodrigo **Neves**

Machine Learning Engineer

rodrigo.f.l.neves@gmail.com | rnev.es | RodrigoNeves95 | rodrigoflneves

"Only a few know how much one must know to know how little one knows"

Summary

Currently working as a contractor to a consultant company as a Machine Learning engineer, with previous experience (4 years) working as a Software/Machine Learning engineer at a start-up with a focus on building predictive solutions for the energy and industry sector. Started working on a pure data scientist position, but steady and slowly started to earn more responsibility and work in different positions and tasks like developing production-ready software, designing and building internal and external dashboards, presenting solutions to clients, deploying and managing workflows over a Kubernetes cluster and working on the frontend and backend of the core product. That led me to learn skills from different fields like machine learning, software and full-stack engineering. This exposure to many areas of the software cycle allows me to understand how they interact and impact each other, which I consider really important in a fast-paced environment like it is the software field nowadays. Nevertheless, I know that I still have a lot to learn and every day it seems that I know even less than the day before. The development pace of new technologies, frameworks and tools can be daunting and terrifying, but you will have to face it as new opportunities and excitement. Above positions and/or titles what I love most about my daily work is to feel challenged and uncomfortable. I consider myself as someone who is eager to learn, loves a good challenge, likes to ask questions, do not like inefficiencies, aims for excellence and is always ready to help.

Work Experience

Daredata Engineering

Lisbon, Portugal

Machine Learning Engineer — *June 2021 - Present*

- Migrated cross-selling models from R into Python, further improving them with new feature engineering and models architectures and implementing distributed hyperparameter optimization on Spark with Hyperopt. Pushed best practices from MLOps perspective to easily operationalize the newly developed models, using Mlflow and Databricks on Azure.
- Geoeonconding system to automatcly calculate available area from street address
- Built and deployed on Google Cloud Platform an infrastructure to deploy and schedule GPU enabled ML models using a Kubernetes cluster and Airflow as scheduler.
- Planned and developed a proof of concept solution for a big utility company to create an automatic and interactive solution to calculate solar potential through satellite images. Implemented a computer vision pipeline using Pytorch and OpenCV to generate all the artifacts and data needed to integrate with the GIS-based systems. As a second phase, a machine learning model was developed to calculate the probability of a costumer buying solar panels, based on the output of the first phase.

Jungle AI

Lisbon, Portugal

Software/Machine Learning Engineer — *May 2018 - June 2021*

- Researched and implemented Jungle's core approach to predictive modelling with an emphasis on recurrent models and time series data implementing from scratch most of the modules using Pytorch.
- Active member of 5+ projects within the energy and heavy industry sector with the main purpose of showing Jungle's predictive solutions. This work would evolve to gather, process and analyze data to apply Jungle's predictive maintenance pipeline and presenting the result to the clients.
- Designed, developed and build internal and production-ready software and deployments. Tasks like ORM definition for production databases, software to process essential high-volume data analysis or the development of a backend infrastructure to manage production deployments using FastAPI, among others.
- Designed and implemented a database architecture (Postgres and TimescaleDB) refactor to allow for a scalable, high-performing and cost-effective solution.
- Built an automated data collection and cleaning workflow (AWS S3 and Prefect) to process daily 2k+ sensors with a millisecond frequency.
- Built a CRUD RESTful API in Node.js for TimescaleDB and PostgresSQL database to be used by Jungle product.
- Implemented, deployed and maintained several high-performing, scalable and distributed workflows using Argo and Kubernetes to different business needs.
- Designed, built and managed internal dashboards and data warehouse solutions using InfluxDB and Grafana to allow for high-volume data analysis.

Jungle AI

Lisbon, Portugal

Master Thesis Internship — *August 2018 - May 2018*

- Had the opportunity to do my master thesis work in a professional environment. The main research goal was to study, implement and test the performance of different deep learning models and architectures applied to time-series problems, specifically in the area of predictive maintenance and forecasting. This allowed me to be an active member of Jungle's daily projects to establish their technology.

Technical Skills

Programming Language	Python, Javascript
Machine Learning Stack	Pytorch, Tensorflow, Scikit-learn, XGBoost, LightGBM, MLFlow
Data Analysis Stack	Pandas, Numpy, Dask, Spark, Scipy, Ray, IPython, Holoviews, Bokeh, Plotly, Matplotlib, OpenCV, Shapely
Workflow Orchestration	Docker, Kubernetes, Argo, Prefect, Airflow
Fullstack Technologies	PostgreSQL, TimescaleDB, InfluxDB, Node.js, Flask, FastAPI, ReactJS, Gatsby, SQLAlchemy
Software development	Git, CI/CD, Unit-testing, Agile, GCP, Azure, Databricks, Distributed Computing

Education

Insituto Superior Tecnico - University of Lisbon *Lisbon, Portugal*

Msc. in Electrical and Computer Engineering *Sept. 2012 - June. 2018*

- Masther thesis - "An overview of deep learning strategies for time-series prediction". Work done in collaboration with Jungle AI.