João Correia

joaodpscorreia@gmail.com | linkedin.com/in/joao-diogo-correia/ | github.com/jdpsc

Machine Learning Engineer with over 3 years of experience in multiple AI applications. Proficient in the entire machine learning life cycle (MLOps), including model management, deployment, and monitoring in production environments. Additionally, I have 2 years of experience in team leadership, mentoring and guiding team members, as well as managing project timelines to ensure successful delivery.

Technical Skills

Languages: Python, SQL, Java, C, C++
AI areas: Deep Learning, Time Series Forecasting, Anomaly Detection, Computer Vision, Natural Language Processing
MLOps: Experiment Tracking, Model Deployment, Model Monitoring, Pipeline Orchestration, Cloud Services (AWS and Azure)
Libraries and Frameworks: Pytorch, Numpy, Pandas, Scikit-Learn, Tensorflow, Mlflow, OpenCV, Hugging Face Platform, Docker, FastAPI, Streamlit, Airflow

Experience

Machine Learning Engineer October 2021 – Present
Jungle AI *Lisbon, Portugal*

- Updated the architecture of predictive maintenance models, as well as optimizing training speeds, allowing the creation of a deep learning foundation model trained on millions of data points, with the end goal of cutting the time and effort of future deployments by around 50%;
- Progressed to team leader after a year (August 2022); Participated on product road map discussions, helping shaping the company strategy; Mentored and guided other team members, ensuring proper technical and personal development;
- Led the design, development and deployment of a power forecasting API service, based on the transformer deep learning architecture, that outperformed competitors in regards to the MAE, launching a new product and creating a new revenue stream for the company; Gathered both business and technical requirements and ensured their alignment across other machine learning engineers, software engineers, commercial team members, and customers.

Data Science Instructor April 2022 – Present
Lisbon Data Science Academy *Lisbon, Portugal*

- Organized two hackathons, preparing materials and providing support to competing teams;
- Reviewed and improved materials for one of the learning units, "Machine Learning for Time Series";
- Organized a deep learning mini-course, presentation and Q&A session.

Scientific Researcher December 2020 – September 2022
ISR - Institute for Systems and Robotics *Lisbon, Portugal*

- Worked on two computer vision projects: the first one action anticipation, the second 3D object detection, both in the context of autonomous driving;
- Published a paper detailing the methodology and analyzing accuracy improvements over other state of the art approaches.

Education

IST - Instituto Superior Técnico Lisbon, Portugal
Integrated Masters (B.Sc and M.Sc) in Electrical and Computer Engineering *September 2016 – October 2021*

- Final grade: 18/20;
- Awarded the University of Lisbon merit scholarship in two academic years;
- Joined Aerotéc, the Aerospace Engineering Students Association, where we designed a computer vision based autonomous landing system for a UAV.

Projects

Deep Learning Course | *Pytorch, Pandas, Numpy, Hugging Face Platform*

- Together with two industry peers, designed a deep learning course with the goal of offering a pragmatic introduction to Deep Learning, including CNNs, RNNs, Transformers, and Pre-trained models;
- Designed an NLP sentiment analysis project to demonstrate practical applications of deep learning;
- Organized a cohort of 8 students, holding weekly sessions to go over student solutions and provide feedback.

Market Price Predictor | *Pytorch, Pandas, Hopsworks, FastAPI, Streamlit, Docker, Airflow, Mlflow, AWS (S3, ECS)*

- Developed a proof of concept application to predict day-ahead market closing prices, primarily focused on exploring MLOps practices;
- Used Hopsworks as a feature store, Mlflow + S3 to manage the model lifecycle, Pytorch + Lightning to define and train models, and Airflow to orchestrate and schedule the multiple tasks - data ingestion, inference and monitoring;
- Developed an API using FastAPI to obtain the predictions, alongside a Streamlit web application for visual inspection;
- Deployed the application using AWS ECS Fargate (Elastic Container Service).

Published Papers (First Author)

Pedestrian Intention Anticipation with Uncertainty Based Decision for Autonomous Driving

- Developed a method for pedestrian action anticipation under uncertainty in autonomous driving, using deep learning and data from multiple sensors;
- Improved anticipation accuracy on two comprehensive datasets by around 6.5% compared to previous works.

Learning Performance Models of Distributed Computer Vision Methods for Decision Making in Detection and Tracking Algorithms in UAVs

- Developed a method for optimizing algorithm selection in UAV-based computer vision tasks, leveraging CNNs and classic machine learning methods to estimate algorithm performance based on image properties;
- Demonstrated superior performance compared to existing methods, reducing computing time while maintaining the same accuracy.

Kaggle Challenges

Automated Essay Scoring | *Pandas, Scikit-Learn, Hugging Face Platform, LightGBM*

- Goal of this code competition was to train a model that scores student essays on a 1 to 6 scale (rounded);
- Experimented with two approaches: directly fine-tuning a single LLM (DeBERTa); and combining the embeddings of multiple LLMs with a LightGBM model, obtaining a best score of 0.81 (quadratic weighted kappa).

Certifications

Cisco Certified Network Associate | *Routing, switching, VLANs, subnetting, TCP/IP, IPv4/IPv6, NAT, ACLs*

Azure Data Fundamentals | *Azure data services, storage, processing and analytics*

Azure Data Scientist Associate | *Azure machine learning services, model deployment*