

João Pereira

Amsterdam, The Netherlands

🌐 www.joao-pereira.pt | 📞 (+31) 06-19792025 | ✉ mail@joao-pereira.pt | 📱 [jpcpereira](#) | 📺 [jpcpereira](#)

Education

EngD in Data Science

Eindhoven University of Technology

Eindhoven, The Netherlands

2019-2021

MSc & BSc in Electrical and Computer Engineering

Instituto Superior Técnico

Lisbon, Portugal

2013-2018

Exchange Student

Université Catholique de Louvain

Louvain-la-Neuve, Belgium

2016

Experience

Senior Data Scientist

adidas

Amsterdam, The Netherlands

Since Jan. 2023

As Tech Lead of two ML products, my main responsibilities have been:

- Developing production-ready machine learning models that forecast demand for thousands of adidas products and ensure their availability in .com and apps.
- Building and maintaining highly scalable, robust, cloud-native machine learning pipelines deployed on AWS (SageMaker Pipelines that orchestrate training, inference, and processing steps; EMR; Step and Lambda functions).
- Integrating MLOps components into the forecasting products (e.g., automated model monitoring).
- Developing CI/CD pipelines (Jenkins), unit tests (PyTest), code quality checks (PyLint, Black), integration tests.
- Optimizing big data processing workflows (PySpark).
- Collaborating closely with adidas' Global Buying, Planning, and Trading teams to promote adoption of our products and ensure their impact.
- Rolling out ML products to new users and markets.
- Streamlining data science tools and way of working.
- Hiring data science talent.

Data Scientist

adidas

Amsterdam, The Netherlands

March 2021 to Dec. 2022

- Developed a neural sequence model for in-season demand forecasting (PyTorch).
- Developed a gradient boosting model for pre-season demand forecasting (Scikit-learn).
- Deployed the two ML models using SageMaker Pipelines.
- Rolled-out the two ML products to the EU market.
- Led cutting-edge project on Graph Neural Networks with the AWS Machine Learning Solutions Lab.
- Supported evaluation of the ML Platforms.

Data Scientist

FIOD - Belastingdienst

Utrecht, The Netherlands

Jan. 2020 - Feb. 2021

In my EngD final project with the data science team of the Dutch Tax Authorities (CoDE), my tasks were:

- Developed an object detection model based on a YOLO architecture that detects 40 classes (TensorFlow).
- Developed an image classification model for cars (200 classes).
- Deployed the models for serving via RESTful APIs using Flask and Docker.
- Built a web application (FIOD Image Intelligence) that enables loading images, triggering model inference, and displaying the results (JavaScript, NGINX, D3.JS, MapBox).

EngD Candidate in Data Science

Eindhoven University of Technology

Eindhoven, The Netherlands

Jan. 2019 - Feb. 2021

As a doctoral candidate in the EngD Data Science, I have:

- Worked on seven data science projects with multiple companies (ASML 2x, Van Lanschot, TE Connectivity, Heijmans, FIOD, Municipality of Den Bosch). Check my website for project details.
- Coached professionals in the Data Science for Health Program
- Organized the Introduction to Data Science elective course at TU Eindhoven.

Student Researcher

Institute for Systems and Robotics

Lisbon, Portugal
Jan. 2018 - Dec. 2018

In the final year of my Masters, I was affiliated with the Signal and Image Processing group of ISR:

- My research focused on deep learning for anomaly detection in time series data. I have worked extensively with variational autoencoders, recurrent neural networks, and attention mechanisms. The proposed approach was unsupervised, making it suitable for applications where obtaining labels is expensive or time-consuming (fraud detection, medical diagnosis, fault detection, ...).

Intern (2x)

{Lisbon, Coimbra}, Portugal

EDP Group

Summer of 2016 and 2017

In my internships at EDP, I have:

- Built a neural network-based model to forecast the power load in hundreds of electrical substations within the Planning Department of EDP.
- Worked on the design of medium/low voltage networks within the Network Studies team.

Skills

| | |
|-----------------------------------|--|
| Languages | Python, Matlab, JavaScript, C/C++, R, LaTeX |
| Cloud | AWS (SageMaker {Pipelines, Studio}, Step and Lambda functions, Glue, Athena, S3, ECR, IAM) |
| Machine Learning | Anomaly detection (my specialty), regression, classification, unsupervised learning |
| Deep Learning | CNNs, RNNs, Seq2Seq, attention, generative modeling (VAEs), reinforcement learning |
| Libraries & Frameworks | TensorFlow, Flask, PyTorch, OpenCV, Scikit-learn |
| Big Data Processing | Spark |
| Frontend | JavaScript, Bootstrap, jQuery, HTML, CSS, D3.js |
| Soft | A lot of enthusiasm and energy for solving challenging problems. |

Selected Publications

Please check [my website](#) or [my Scholar profile](#) for a full publication list.

Fast and Scalable Hyperparameter Tuning and Cross-validation in AWS SageMaker

Medium article

March 2023

Unsupervised Anomaly Detection in Energy Time Series Data Using Variational Recurrent Autoencoders with Attention

Orlando, USA

2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)

Dec. 2018

Learning Representations from Healthcare Time Series Data for Unsupervised Anomaly Detection

Kyoto, Japan



2019 IEEE International Conference on Big Data and Smart Computing (BigComp)

2019

Trainings

Manager Development Experience | 6-month program | adidas | Virtual
AWS Distributed Training and MLOps | 2-day Workshop | AWS | London, UK
AWS SageMaker Inference (Batch and Real-time) | 1-day Workshop | AWS | Sept. 2022 | Virtual
Spark Processing in AWS SageMaker | 1-day Workshop | AWS | Sept. 2022 | Virtual

Talks

Fast and scalable hyperparameter tuning and cross-validation in SageMaker | Amazon Web Services | March 2023 
Project Holy: demand forecasting with graph neural networks | adidas | Jan. 2023
Enabling trading excellence via Fortuna's pre-season forecasts | adidas | July 2022
In-season demand forecasting with DeepAR | adidas | Sept. 2021
Anomaly detection: overview, strategy, and use cases | adidas | May 2021
Anomaly detection with variational autoencoders | Deep Learning Sessions Lisbon | May 2020 
On Deep Learning | EngD Data Science | March 2020 