

# David Amat

Senior Data Scientist  
(Glovo)

CAT | ENG | SPA

Location: Barcelona, Spain  
Date of birth: 20-04-1994

Number: +(34)-608526629  
daolondrizdaolondriz@gmail.com

in: david-amat-olondriz/  
github: DavidAmat

Full Stack Data Scientist with 6+ years of experience in end-to-end ML development, specializing in recommender systems, deep learning, NLP, computer vision, and AI/LLM technologies. Currently leading projects at Glovo, driving improvements in ML platform lifecycle, feature engineering, and real-time API serving with a focus on scalable, product-driven solutions. Proven track record of collaborating cross-functionally with Product and Backend teams, delivering impactful data-driven products. Constantly advancing skills in AI/ML system design and staying updated with state-of-the-art technologies.

## Work experience

### Glovo | Senior Data Scientist

Barcelona, Spain | 07.2020 - Current

- Led the development and optimization of a Stock Prediction Model (SPM) for Groceries, significantly reducing Product Not Available (PNA) metrics by 10-14%, resulting in ~€2M annual savings.
- Engineered a new restaurant recommender system, transitioning from collaborative filtering (SVD) to an advanced embedding-based model using Word2Vec, leading to a 1% absolute increase in CVR. Integrated the system with a low-latency API utilizing a Feature Store, enabling real-time recommendations. A/B test experiments showed +0.8M incremental orders / year.
- Ideation and productization of Multi-Armed Bandit framework to tailor recommender widgets for each user user, using a custom-made approach iterated from typical Thompson Sampling approach (in test).
- First prize in an internal Hackathon fine-tuning a Stable Diffusion model to generate dish images from text.
- Benchmarking CLIP model on our released Foodi-ML multi-modal dataset
- Gained ML Engineering skills mastering our in-house ML Platform architecture, including project setup with Poetry, scalable trainings in Argo Workflows under K8s, API inferences serving with Flask, Dockerization and ECR management, full CI/CD lifecycle with Argo CD, GitOps, Jenkins, Argo Rollouts, etc. Besides, I also build declarative Data Products running in Airflow, linking them to Feature Store for production low-latency retrieval.
- Experience in A/B test analysis, including design on our Experimentation Platform, metric selection, MDE analysis, and result evaluation using chi-squared tests, z-tests, and p-value corrections with Python libraries

### Banc Sabadell | Data Scientist

Sant Cugat del Vallès, Spain | 11.2018 - 07.2020

- Young Talent program focused on Data Science within the Bank (6 months rotation to different areas).
- Developed a predictive model optimizing channel selection for debt recovery, leveraging decision trees to identify optimal communication strategies, which improved recovery rates by 5%
- Collaborated on a spending categorization model that automatically classifies transactions into broad categories
- Developed a tree-based fraud detection model to flag suspicious transactions and reduce False alarms, reducing on average ~60-70% of False positive alarms, reducing analysts workload on analyzing and reporting them.

### EY | Junior Consultant at Life-Science team

Barcelona, Spain | 07.2018 - 11.2018

### IRB/IBEC | MSc Internships

Barcelona, Spain | 06.2017 - 07.2018

- [IRB] Developed a target prediction tool using Similarity Ensemble Approach (SEA) and Siamese NN, integrating bioactivity data and gene expression profiles to enhance drug-target interaction predictions.
- [IBEC] Used finite-element methods to compute visco-elastic forces for biological tissues .

## Education

IMF School | Master's Degree in Business Analytics & Big Data

Grade: 9.8

Madrid, Spain | 05.2019 - 10.2020

UPF/UB | Master's Degree in Bioinformatics for Health Sciences

Grade: 9.1

Barcelona, Spain | 09.2016 - 06.2018

UPC | Bachelor's Degree in Engineering Physics

Grade: 7.6

Barcelona, Spain | 09.2012 - 06.2016

## Professional Development

---

Continuously developing expertise in AI/ML through courses on LLMOps, NLP, and Big Data technologies, including hands-on experience with cloud platforms like AWS.

### MOOCs:

**LLMOps & ML Deployment:** *Udemy* | *Current*  
**NLP Specialization** *Coursera* | *01-2022*  
**Deployment of ML models** *Udemy* | *12-2020*  
**Neo4J Introduction** *Neo4j* | *06-2020*  
**Plotly/Dash dashboards** *Udemy* | *02-2020*  
**Big Data Specialization** *Coursera* | *03-2019*  
**Mathematical Biostatistics** *Coursera* | *05-2021*

### Books:

**Hands-On Generative AI** *O'Reilly* | *Current*  
**Hands-On Graph NN** *O'Reilly* | *01-2023*  
**NLP with Transformers** *O'Reilly* | *07-2022*  
**Applied recsys with Python** *O'Reilly* | *01-2022*  
**Deep RL Hands-On** *O'Reilly* | *12-2021*  
**Hands-On Explainable AI (XAI)** *O'Reilly* | *07-2021*  
**Bayesian Analysis with Python** *O'Reilly* | *03-2021*

## Publications

---

- FooDI-ML: a large multi-language dataset of food, drinks and groceries images and descriptions | *arXiv* - 2021
- Extending the small-molecule similarity principle to all levels of biology | *Nat Biotechnol* - 2020
- Computation of forces from deformed visco-elastic biological tissues | *Inverse Problems* - 2018
- Pendulum dynamics in an amusement park | *European Journal of Physics* - 2017

## Projects

---

- **2048:** developed in Python a clone of the 2048 game and used REINFORCE to play the game optimally
- **Foodi-ml:** training image-text alignment model (ADAPT) for FooDI-ML dataset
- **Soccer-Stars Django Webapp:** clone the mobile game Soccer Stars with a React app using Django as Backend to compute collisions and trajectories
- **Fitchek AI:** build a Django Backend connecting to OpenAI API to ask GPT-4o for ratings on your outfit image
- **Portfolio:** building my own web app portfolio in Next.js / React / Tailwind CSS, etc. (*In Progress*)
- **PDF Search LLM RAG:** building a local PDF summarizer with Sentence Transformers, OpenSearch as Vector Database, Mistral fine-tuning with LoRA on question/answers dataset generated by GPT-4o, and serving it as a Streamlit app that uses RAG from the embedded PDFs. (*In Progress*)
- **Build your own MLP:** building a Medium post to guide on the set up of a local K8s cluster with K3d, with Argo installed to leverage your computer as a compute server using Argo Workflows and Rollouts for the API (*In Progress*)

## Skills

---

**Programming Languages:** Python | R | SQL

**Web/Backend Development:** Django | Flask | React

**Machine Learning & AI:** NLP | Computer Vision | PyTorch | HuggingFace | Transformers | Generative AI | Multi-Armed Bandit | REINFORCE | Collaborative Filtering | Word2Vec | Graph NN | Bayesian Inference | RAG

**Data Engineering & MLOps:** Feature Store | Trino SQL | Airflow | Docker | Kubernetes | Argo | Jenkins | AWS | MLOps | Unit/Integration Testing | CI/CD

**Tools & Platforms:** Git | Grafana | Datadog | OpenAI API

## Interests

---

Playing the piano (I love *Chopin*), playing sports video games like *FIFA*, I've been a tennis coach for 4 years as a teenager, now I compete in local competitions.