TOMAS PEREIRA DE VASCONCELOS

"The good life is inspired by love and guided by knowledge"
— Bertrand Russell

tomasvasconcelos1@gmail.com   +31 6 86 11 22 17   linkedin.com/in/tpvasconcelos   tomasvasconcelos.com

🚀 Ambitious

- An (aspiring) grandmaster of the Data Science Lifecycle.

⚡ Professional experience

- I have developed into a **mature, reliable, and resourceful engineer**, equipped to solve complex machine learning and software engineering problems.
- From machine learning engineer to solutions architect, data engineer, and data scientist... I have worn many hats while helping teams and companies **deliver value-driven solutions** to predictive analytics and data science problems.
- I have hands-on experience in **solving real business problems** using techniques from time-series forecasting, recommender systems, reinforcement learning, Bayesian inference, and many others.

📚 Academic experience

- In 2017, I graduated from Royal Holloway University of London with a **First Class Honours degree in Physics**, followed by a **Mathematics graduate degree** at King's College London (KCL).
- During these years, I also developed some research experience in computational fluid dynamics and contributed to **a publication** in the field of theoretical statistical mechanics.

🐝 SEO Buzzwords

- **Python | SQL | Docker | Big Data | Spark | AWS | Kubernetes | MLOps | CI/CD | Agile | Microservices**

Experience

Beat / Machine Learning Software Engineer / Amsterdam, Netherlands / Jun 2021 - Present · 10 months

Beat is the fastest-growing ride-hailing service in Latin America. My mission is to enable and accelerate Beat's sustainable growth by developing and applying Data Science and Machine Learning solutions to detect and prevent fraudulent actions, reduce financial losses and abuses, and ensure a safe environment for all users of our platform. All of this while keeping the business metrics healthy and the company growing!

🔧 Feature Store Deployment (w/ Feast)

Led the development and implementation of a Feature Store solution to serve all ML and Analytics teams at Beat. Started by evaluating managed solutions such as Databricks, and SageMaker, conducting a two-week trial of Tecton, and finally settling for Feast as the registry and serving layers. Deployed the solution end-to-end on our Kubernetes cluster using our data-lake and Trino as the offline store, ElastiCache as the online store, and a low-latency gRPC Java service. Worked closely with Feast's dev team by providing feedback and contributions to the open-source project. Currently working on standardising batch transformations in Fraud.

📇 Fraud Prevention

Led the productionisation of a fraud prevention system by (1) designing the solution's architecture, (2) validating the assumptions and performance estimations of Data Scientists, (3) developing and deploying an ML (Argo) workflow that computes risk scores and pushes them to a Kafka topic, to be ingested to a fast-store (Aerospike) by a BE system, (4) configured monitoring for the infrastructure and application with Prometheus and Grafana, and (5) designing and running an online controlled experiment.

🚕 Fraud Prediction

Worked with another Data Scientist on a fraud prediction problem where the labels were sparse, biased, noisy, and mostly positive and unlabelled (PU). We considered and explored several modeling options from naive binary classification with gradient boosting models to more robust active-learning approaches. The latest implementation lies somewhere in the middle, based on an iterative semi-supervised learning solution that yields twice as many fraud cases as the previous system while keeping precision extremely high (as per the pre-defined acceptance criteria).

Tiqets / Machine Learning Engineer / Amsterdam, Netherlands / Jan 2019 - May 2021 · 2 years 5 months

As a Machine Learning Engineer, I was part of Tiqets' core Data Team. I worked closely with data and business analysts, data engineers, as well as product owners and management team. I applied software development, data analytics, and machine learning to scale and operationalise statistical models and make the whole organisation more data-driven.

📈 Time Series Forecasting

Operationalised and automated sales forecasting at Tiqets by deploying a generalised ML workflow to run pre-processing, model selection, evaluation, and periodic updating of forecasted values for multiple and varying business contexts. Each task was distributed across an array of Celery workers deployed on our Kubernetes cluster. Forecasted values and model metadata was pushed to our DWH and visualised by business stakeholders using our BI tool. We used DataDog for both application and infrastructure monitoring. For the most part, LSTM neural networks did not yield better results than classical approaches such as ARIMA and Exponential Smoothing, while an optimised Prophet model performed best for our type of data.

🎥 Recommender System

Improved recommendations across our platform by taking user-preference and item-similarity into account. Also, enriched popularity-based heuristics for cold-start instances by taking distance and recency into account. A curated version of an Amazon Personalize model now powers a great part of recommendations at Tiqets. To help the Data Team iterate faster and with greater confidence on new versions of the recommender models, we also implemented a time-dependent offline evaluation pipeline for recommender systems, curated for the e-commerce setting.

🏆 Learning to Rank with Reinforcement Learning

To optimise any ranked list on our platform, I initially framed this task as a Supervised Machine-Learned Ranking problem, which involved comparing the predicted ranking to some *ideal* ranking using a metric such as nDCG and fitting a simple heuristic to this. To improve on this iteration, I implemented Bayesian Reinforcement Learning-to-Rank bandit strategies (e.g. Thompson sampling), which use *explore and exploit* to continuously learn and improve the rankings. As new data is collected through our event-pipeline, we use Airflow to frequently update item rankings on our platform (e.g. by re-sampling from an updated posterior distribution).

👨🏽‍💼 Leadership / Management / Business Skills / Soft Skills

• Helped management (CEO, COO, and CTO) and product owners in brainstorming sessions, defining company OKRs and team KPIs, and led meaningful reporting initiatives and important ad-hoc analysis.

• The team grew from 3 to 15 people since I joined. I helped with the recruiting, assessing, and interviewing of candidates (including the current Head of Data) as well as organising and attending career fairs.

• Provided guidance and supervision to two University students working on their Master's Thesis. Both students produced valuable projects for Tiqets and finished with outstanding grades (8/10 and 8.5/10).

Accelogress / Software Engineer (Machine Learning) / Gildford, UK / Jun 2016 - Mar 2018 · 1 year 10 months

As a Machine Learning Engineer, I worked directly with the CEO and Lead Developer at Accelogress Ltd, a software consultancy company developing solutions using machine learning and API technologies. For the Save-a-Space project, I led the development of (1) the forecasting engine for car park availability using a robust machine-learning framework, and (2) a pre-processing layer and REST API to expose historical, real-time, and forecasted availability for multiple car parks around the UK, to our mobile app and web dashboards.

Technologies: Docker, nginx, Gunicorn, Django REST Framework, JavaScript, AngularJS, scikit-learn, MySQL

Research / Publications

📄 *"The second virial coefficient of bounded Mie potentials"* - *The Journal of Chemical Physics* 147-214504 (2017)

This work is concerned with, for the first time, the derivation of analytic expressions for the second virial coefficient of bounded Mie potentials. The convergence properties of the series expansions were also considered, as well as considerations of parameters that give rise to thermodynamic stability.

📖 *"Soft-Core Lennard-Jones model"* - BSc Physics Thesis (82%)

My research was focused on the mathematical properties of a special potential model used to describe the interaction between particles in colloid and polymer systems. Official feedback from thesis supervisors:

– *"Tomas observed and brought out [...] a number of very interesting results, of a caliber which would reflect well on a researcher of many more years experience."*
– *"The series expansion and MD modeling aspects of the project were definite steps forward in our understanding of the behavior of this system."*
– *"[Tomas] derived single-handedly a series expansion in terms of orthogonal polynomials, which is a new approach as far as I am aware."*

Education

GradD Mathematics / King's College London, UK / Sep 2017 - Aug 2018

- Developed a robust mathematical foundation for further studies and research in theoretical physics.
- Refined my mathematical intuition to interpret and build models for simple and complex systems.
- The skills acquired are highly transferable to other fields of applied mathematics and statistics.

BSc Physics / Royal Holloway, Uni. of London, UK / Sep 2013 - Jun 2017 / First Class Honours

- Developed the intuition to identify relevant laws and principles and apply appropriate mathematical tools and approximations when working on complex problems.
- Exposed to advanced experimental methods in collection of experimental data, evaluating its significance, drawing relevant conclusions, and calculating the significant statistics.
- Developed several scientific computing projects with a focus on data analysis and fluid dynamics.
- Experience in communicating experimental results and producing clear and accurate reports.

High School (Exchange Year) / River Ridge High School, USA / Aug 2011 - May 2012 / 93% GPA

- Took part in a one-year exchange program on my last year of high-school.
- Top grades for mathematics, physics, and chemistry with scores ranging between 95% and 100%.

High School (Science Track) / Salesianos de Manique, Portugal / Sep 2008 - Jun 2011 / 80% GPA

- Ranked in the 97th percentile for the Physics & Chemistry national exam.