

Florian J. B. Leitner

C/ XXX

E-28005 Madrid

☎ +34-XXX-XXX-XXX

🌐 datacatalytics.com

🌐 fnl.es

✉ leitner@datacatalytics.com

✉ me@fnl.es

• linkedin.com/in/florianleitner

• github.com/fnl

• stackexchange.com/users/fnl

• scholar.google.com/citations?user=EkXLlysAAAAJ

Data Scientist / ML Engineer

Profile

Engineering manager for machine learning and data science teams.

- coaching mixed teams of up to 25 engineers, researchers, and analysts
- creating analytical NLP systems for information extraction and text clustering
- design and ops of semantic search, percolation, and content-based recommendation engines
- data engineering of low-latency stream and high-volume batch processing systems
- core values: transparency, curiosity, accountability, compassion, and agile communication
- domain expertise in biotech (omics & oncology) and finance (ECM/DCM/FICC)

Biggest career achievements: Being the CTO of a financial news entity that successfully exited, receiving a PhD in molecular biology (with distinction), publishing a Nature Biotechnol. first author paper, and founding a data science consultancy.

Career

Head Of Data Science, ION Analytics, New York, USA – 2021-Present

- ▶ Working within the executive team (CPO & CTO) of the company's Product function on formulating our strategic objectives and resolving day-to-day business matters.
- ▶ Working with data scientists and engineers on defining the strategic objectives for the company's augmented intelligence systems in accordance to the executive team's vision.
- ▶ Coaching data scientists and engineers on technical issues and aspirational targets to align their contributions to the company's best practices and mission while building a world-class, high-performance machine learning team & culture.
- ▶ Architecting state-of-the-art machine learning operations to enhance stakeholder (analysts, stringers, journalists, editors, lawyers) productivity and increase data coverage & precision.

Engineering Manager, ION Analytics, New York, USA – 2019-2021

- ▶ Leading the development of a search engine for structured financial data by extending the capabilities of Elasticsearch and enhancing the UX with natural language query capabilities, thereby engaging some of the largest investment companies in a brand new financial product.
- ▶ Enhancing an in-house real-time content recommendation system by integrating Elasticsearch to enable the delivery of document body snippets (highlights) and reducing the engine's overall memory footprint and cluster size, saving over \$100k YOY in operational costs.
- ▶ Coaching a team of analysts and engineers to scale up the reference data system towards the goal of adding two orders of magnitude more entities.
- ▶ Managing the implementation of a UI-based text classifier training framework to make the process of adding new topic tagging capabilities independent of the data science team.

- ▶ Coordinating a TechOps team that kept expanding the CI pipeline capabilities and automated tedious DevOps tasks to continually make the developer team more productive.

Chief Technology Officer, Selerity Inc., New York, USA – 2018-2019

- ▶ Led a globally distributed team of 25 through a successful acquisition.
- ▶ Continually improving the quality of a proprietary recommendation system for public and private content with a team of data scientists by evolving the in-house NLP libraries for news-like articles that then are percolated to employees of some of the world's largest banks.
- ▶ Overseeing the development of a credit trader chat analytics engine to predict trade partners, detect missed opportunities, and provide trade analytics that was sold to some of the biggest intra- and inter-bank Sales & Trader Platforms.

Faculty Member, IE University, Madrid, Spain – 2018-Present

- ▶ Teaching core concepts in AI and Machine Learning to MBA students and improving the class using the students' evaluation feedback.

Senior Full-stack Data Scientist, Selerity Inc., New York, USA – 2015-2018

- ▶ Evolving the in-house open information extraction and text classification systems for financial instruments, legal entities, and business concepts/topics by increasing systems performance against continuously annotated data generated from the QA process.
- ▶ Designing unsupervised text clustering algorithms for news and proprietary financial documents to detect correlations and trends by iteratively improving analyst reviews of cluster coherence.
- ▶ Domain-driven design of entities and aggregates needed for a trader chat analytics engine by collaborating with credit desk experts to establish all requirements and architect the data structures that can encode the trader workflow.
- ▶ Tuning and maintaining Selerity's in-house message streaming platform (similar to NATS or Akka) by reducing end-to-end latency and resource use (CPU load & memory usage).

Full-stack Data Science Consultant, Data Catalytics, Madrid, Spain – 2015-2018

- ▶ Worked remotely as a part-time international consultant on a variety of NLP-related projects, such as extracting structured data from CVs or patents, delivering the respective project goals as defined at contract onset.

Instructor, Adv. Statistics and Data Mining Summer School, Madrid, Spain – 2014-Present

- ▶ Teaching the NLP class component of a prize-winning summer school (5x3h lectures) with the objective of, due to MOOCs, simply maintaining the number of participants, while improving class quality through participant feedback.

Juan de la Cierva fellow, Univ. Politecnica, Madrid, Spain – 2014-2015

- ▶ Researched probabilistic graphical models and developed text mining systems in neurobiology with the dual goal of facilitating knowledge acquisition for neurobiology researchers and publishing novel research in the field.
- ▶ Published a research article while teaching and mentoring PhD and master students.
- ▶ Invited speaker at research conferences and events.

Post-doctoral Fellow, Span. Nat. Cancer Research Centre, Madrid, Spain – 2012-2014

- ▶ Developed a tool to extract transcription factor-gene interactions with the experimental evidence for the interaction from scientific publications and delivered on the objective of providing systems biologists with an autogenerated mammalian transcriptome.
- ▶ Published various research articles and presented at conferences globally.

PhD Student, Span. Nat. Cancer Research Centre, Madrid, Spain – 2006-2012

- ▶ Led a remote, global team of investigators across twelve research groups to develop the world's first distributed BioNLP engine with the goal of using it for online community evaluations, proving both that a distributed system is feasible and advances the state-of-the-art.
- ▶ Completed a PhD by developing a metric that evaluates both set and ranking quality as a single scalar (F-measured Average Precision), demonstrating that it leads to superior results than bootstrapping with more common performance metrics to tune an ensemble meta-learner.
- ▶ Organised one and co-organised two community evaluations (BioCreative), and developed the community website that is used to organise the events and manage the evaluations.
- ▶ Collaborated in the development of toxicology mining, automated ontology construction, and gene and chemical entity extraction systems.
- ▶ Published a dozen research articles, incl. a Nature Biotechnol. first-author paper, and presented posters & talks at conferences.

Master Student, Institute for Molecular Pathology, Vienna, Austria – 2003-2006

- ▶ Extended a molecular sequence analysis suite used to investigate post-translational modifications with new new sequence and protein structure repositories requested by researchers working with the suite, by creating parsers and designing DB schemas.
- ▶ Created UI views for the sequence and protein structure repositories and their meta-data using a LISP DSL to auto-generate a Struts-based web presentation layer.
- ▶ Made the repositories and their meta-data searchable by indexing them with Lucene, and designed & developed the information retrieval user interface.
- ▶ Co-authored on a poster and a conference publication.

Consultant, CSO Lanifex, Vienna, Austria – 2003-2003

- ▶ Embedded a Linux kernel in (freeS/WAN-based) VPN thin clients and customised Mozilla to deploy it as the thin client's UI to (only) show the remote desktop, for the Austrian National Bank.
- ▶ Database schema development (MySQL) for an internal, networked PIM system used by the U.N. headquarters in Vienna.

Application Developer, HITS, Heidelberg, Vienna – 2002-2003

- ▶ Developed an XML-based annotation system for protein structures as a plug-in for the ProSAT 3D viewer.
- ▶ Set up a Pythonic workbench and environment for structural biology and researched the design neural networks.
- ▶ Co-authored my first research article.

Software Developer, Novartis, Vienna, Austria – 2001-2002

▶ Developed a web-accessible protein-annotation and sequence analytics pipeline for the structural biology research team.

Production & Sales Manager, Pro-Design, Innsbruck, Austria - 1997-2000

▶ Organised the international production (China, Israel, South Korea) of paragliders and kites, and managed a world-wide sales team.

▶ Organised marketing (sports) events and advertisement campaigns.

▶ Collaborated on the in-house production/warehouse management tool with the co-founder.

Assistant Manager, Leitner Siebdruck, Innsbruck, Austria – 1996-1997

▶ Product management and sales; Managed a small team of printers and designers.

▶ Making offers and bids on tenders for commercial signs, markings, and advertisement.

▶ IT operations assistance (maintenance of Apple OS and MS Windows systems).

Education

PhD, Autonomous University of Madrid, 2012

Molecular Biology, Bioinformatics, BioNLP

Business Skills

Team coaching & culture building, trust building & leadership, lean development using the Agile & Kanban methodologies, customer-service and -value orientation, quality assurance, requirements formulation, business policy design, public speaking and writing, financial markets & biotechnology expertise, remote-first working.

Technical Skills

▶ Machine learning frameworks and libraries (TF, PyTorch, SK-Learn, Smile, and *many* more)
▶ Machine learning operations (MLOps) tooling (dvc/cml, eli5, Kibana, MLflow, SageMaker)
▶ DevOps tooling (Ansible, Docker, Graphite, Grafana, Icinga, Jenkins, Rundeck, Travis)
▶ Agile development using XP, T/BDD and DDD (paring, feedback, JUnit, pytest, UML, EER, etc.)
▶ Pub/sub messaging & stream processing systems (in-house, Avro, Akka, Kafka, NATS, Spark)
▶ Search engines (in-house, Lucene/Solr/ElasticSearch, Sphinx, Vespa, Xaipan)
▶ SQL & NoSQL data-stores (Cassandra, CouchDB, MongoDB, MySQL, OrientDB, Postgres)
▶ Web/API development (GraphQL, nginx, Node, Postman, Swagger/OpenAPI, Twisted, vert.x)
▶ Developer tooling ([C]Make, gdb, Git, Grunt, Maven, perf, psutil, sbt, Valgrind, Wireshark, etc.)
▶ Ubuntu, Debian, RedHat & Apple OSX ops experience
▶ AWS Cloud computing (incl. SageMaker); some minor GCP and Azure exposure

Prog. Languages

Expert: **Python, Java**

→ ready to code and teach others, with at least a decade of professional experience

Proficient: **C/C++, Cython, Go, JavaScript**

→ ready to code, but not an expert; at most 5y of accumulated professional experience

Past prof. use: Common Lisp, Erlang/OTP, Perl, Ruby, R, Scala, Visual Basic

→ can parse, but would need a warm-up phase to code in efficiently again; 1-5y of exp.

Basic knowl.: **Haskell, Rust, TypeScript**

→ familiarity, but no professional experience (yet)

Nat. Languages

German, English, Spanish

All fluent.