

Florian J. B. Leitner

C/XXX
E-28005 Madrid

T +34-XXX

W1 datacatalytics.com

W2 fnl.es

E1 leitner@datacatalytics.com

E2 XXX

L linkedin.com/in/florianleitner

G github.com/fnl

Full-stack Data Scientist

Profile

Machine learning research and operations leader/consultant.

Strong domain expertise in biotech (genetics & oncology) and finance (DCM/fixed income). Managing world-wide distributed teams of engineers and researchers. Specialist for open information extraction from, and clustering of, unstructured data. Expert in developing and operating search/recommendation engines. Experienced remote working enthusiast.

Career achievements summary: Being the CTO of a financial news entity that successfully exited, getting a PhD in molecular biology, publishing a Nat. Biotechnol. first author paper, and founding a data science consultancy.

Career

CTO, Selerity Inc., New York, USA - 2018-Present

Leading a globally distributed team of 25 software engineers, data scientists, and QA analysts through a successful acquisition (by the ION Group). Shepherding data scientists in the implementation of state-of-the-art NLP systems (Content Analytics) used to assist ECM and DCM financial professionals. Leading the development of a credit trader chat analytics engine (PCE on IB, Symphony) for predicting trade partners, detecting missed opportunities, and chat analytics. Making Selerity's real-time financial information recommendation engine more efficient (Selerity Context). Managing the scaling-up of Selerity's reference data system. Directing the implementation of a workflow framework to accelerate the data science pipeline. Managing a TechOps team that expanded Selerity's CI pipeline capabilities. Coordinated the analyst QA and reference data team. Mostly remote from Madrid, working in Java, Python (and Scala, Go & C/C++). Cloud: AWS, GCP.

Faculty member, IE University, Madrid, Spain - 2018-Present

Teaching core concepts in AI and Machine Learning to MBA students.

Senior Fullstack Data Scientist, Selerity Inc., New York, USA - 2015-2018

Development of open information extraction and text classification systems for financial instruments, legal entities, and business concepts/topics; Designing unsupervised text clustering algorithms processing news and proprietary financial documents to detect correlations and trends (Selerity Analytics). Designing the data model for a trader chat analytics engine with the goal of enhancing credit desk workflows (Private Context Engine). Tuning and maintaining Selerity's in-house, state-free, low-latency streaming platform. All remotely from Madrid, working in Java, Python (and Scala, C++ & R). Cloud: AWS.

Fullstack Data Science Consultant, Data Catalytics, Madrid, Spain - 2015-2018

Working remotely as a part-time international consultant on a variety of NLP-related projects, such as transforming CVs (PDFs) into structured data, extracting data from publications/patents, etc., with the objective of assisting clients in reaching their particular business goals. Work performed in Scala, C++, and Python. All work done from my Madrid offices.

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

Teaching the NLP class at a prize-winning summer school (5x3h lectures).

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

Researching probabilistic graphical models and developing text mining systems in neurobiology in Scala, C++, R, and Python. Publication of a statistics research article. Teaching and instructing PhD and master students. Invited speaker at research conferences and events.

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

Developing a system to extract the mammalian transcriptome from descriptions in MEDLINE abstracts, associating the TF-gene interactions with their experimental evidence in Scala, C++, and Python. Publication of various research articles. Speaking at research conferences.

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

Remotely leading a global team of researchers to develop the world's first distributed BioNLP engine in Python and Erlang (BioCreative Meta-Server). Completing a PhD by developing a technique that evaluates both set and ranking quality (F-measured Average Precision), demonstrating that it produces superior results when tuning an ensemble meta-learner. Assisting oncologists in their daily research by fulfilling their information extraction needs using Python, R, and Scala. Publication of a dozen research articles, incl. a Nature Biotechnol. first-author paper, presenting posters and speaking at conferences.

Master student, Institute for Molecular Pathology, Vienna, Austria - 2003-2006

Implementing a molecular sequence repository to analyse post-translational modifications using Postgres (Annotator). Creating Perl-based parsers and designing interfaces for sequence meta-data management via a LISP DSL that generated a Struts-based web presentation layer. Distributed indexing of the sequence and meta-data repositories with Lucene and designing an information retrieval interface for Struts in Java. Publishing a first research article.

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

Tuning and minimising the Linux kernel embedded in (freeS/WAN) VPN thin clients for the Austrian National Bank. Custom modifications of Mozilla to deploy it as the thin client's desktop system. Database development/optimisation (MySQL) for a networked PIM system for the internal U.N. headquarters in Vienna.

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

Developing an XML-based annotation system for protein structures as part of the ProSAT 3D viewer in Java. Setting up a Python workbench and environment for structural biology and researching neural network design in C++.

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

Development of a web-accessible protein-annotation and sequence analytics pipeline in Perl for the structural biology research team (NB: strong contractual NDA).

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

Managing the international production (China, Israel, South Korea) and world-wide sales people for a paraglider and kite producer. Organisation of marketing (sports) events and advertisement campaigns. Collaborating on the in-house production/warehouse management tool (VisualBasic) with the co-founder.

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

Production and sales management. Calculations, offers, and bids on tenders for commercial signs, markings, and advertisements. Managing a team of printers and designers. Internal IT operations assistance (Apple OS and MS Windows).

Education

PhD, Autonomous University of Madrid, 2012

Molecular Biology, Bioinformatics, BioNLP

Business Skills

Team leadership, lean flow management, product management, quality assurance, public speaking and writing, financial markets & biotechnology expertise, remote-first working.

IT Skills

Machine learning operations (dvc, MLflow), deep transfer and representation learning (Tensorflow/Keras, Pytorch), polyglot programming (see below), Linux & OSX, cloud computing (AWS, minor GCP knowledge), pub/sub messaging, microservices & stream processing (in-house, Akka, NATS, Kafka), working with (numerous) SQL databases & NoSQL-stores, search engine development and operations (in-house, Lucene/Solr/ElasticSearch, Vespa, Sphinx, Xaipan).

Prog. Languages

Expert: Python, Java, Scala

Comfortable: Go, Cython, C/C++, R

Rusty: Common Lisp, Perl, Ruby, Erlang/OTP