Emilio Dorigatti

Machine Learning Research Scientist, PhD

https://www.linkedin.com/in/edorigatti/en

emilio.dorigatti@gmail.com

Munich, Germany

Professional with 10 years of experience in data-driven AI/ML applications in both industry and academia. Proficient in the in the entire digital value chain, and familiar with the entire data science workflow, from problem definition, data acquisition and cleaning, feature engineering, exploratory data analysis and visualization, model selection, training, evaluation, and interpretation, until model deployment, monitoring, and maintenance. Passionate about computer science and mathematics, eager to work on cutting-edge AI applications in the pharmaceutical industry, keen to continuously grow, learn and share.

Technical Skills

- **AI/ML:** Deep learning, Chemical language models, Geometric deep learning, Graph neural networks, Semi-supervised learning, Representation learning, Computer vision (CV), Object Detection, NLP, Uncertainty quantification, Clustering, Explainable machine learning, Statistical modeling and forecasting, Discrete optimization, Visualization, Model selection, Time Series Analysis, Mathematics, Computational Biology
- **Technologies:** Python, PyTorch, TensorFlow, TensorBoard, MLFlow, Keras, Scikit-Learn, XGBoost, LightGBM, Pandas, Numpy, PyMC3, Jupyter, Spacy, NLTK, R, Shiny, Dplyr, Git, Jenkins, Docker, Spark, AWS, Linux, SLURM, Bash, Make, PostgreSQL, ElasticSearch, Redis, REST, Flask, Django, Java, Spring, Streamlit, FastAPI

Selected Professional Experience

Postdoctoral Scientist, *Boehringer Ingelheim, Ingelheim, Germany,* 10/2023 – Now
Deep learning research for chemical and analytical development.

- Developed SotA generative models for *de novo* molecular structure elucidation conditioned on experimental measurements, made them available to a team of professional chemists through an user-friendly web interface, thereby improving their productivity;
- Participated in the development of QSAR models enabling predictions in novel, challenging data scenarios;
- Bayesian optimization, design of experiments, computer vision for analytical method development and optimization.

Machine Learning Research Scientist, *Helmholtz Zentrum, Munich, Germany,* 07/2019 – 09/2023
Leveraged deep learning for Computational design of personalized immunotherapies for cancer.

- Conceived, designed, implemented, evaluated, and published novel deep learning and machine learning methods in semi-supervised learning for computer vision and natural language processing;
- Conducted thorough analyses of biological datasets, generating and validating data-driven hypotheses;
- Published 17 articles, led 10, research projects, taught 5 university courses, supervised 18 Master's students.

Data Scientist (working student), *BrightCape B.V., Eindhoven, the Netherlands,* 03/2017 **–** 06/2017
Consulted a major Dutch multinational to to optimize staff allocation and website workflow.

- Reported directly to the CTO on a daily basis;
- Performed customer journey analytics and demand forecasting;
- Acted as technical advisor in executive meetings with customers.

Machine Learning Engineer (working student), *Wikidata, Trento, Italy,* 02/2016 – 06/2016
Extension of a knowledge base with 4.5M new triples involving 12M entities.

- Web scraping from over 50 sources (scrapy);
- Developed a text mining framework leveraging named entity recognition and machine learning.

Data Engineer, *SpazioDati Srl, Trento, Italy,* 04/2014 – 06/2016
Worked on the data processing systems underlying the company flagship product.
- Developed an ETL pipeline based on Spark, SQL and NoSQL databases on AWS cloud, aggregating and processing data for 20M entities every week;
- Developed REST microservices for internal use.

Education

Ph.D. in Statistics, *Ludwig Maximilians Universität, München, Germany,* 07/2019 – 09/2023
- Developed optimization frameworks to design vaccines for cancer
- Studied methodological challenges related to artificial intelligence in precision medicine and drug discovery.
- Researched unify deep neural networks with statistical regression models to analyze clinical data.
- Additional courses in Project Management and Time Management.

Double Master's Degree in Data Science, *European Institute of Innovation and Technology, 08/2016 – 06/2018*
- Specialization in Distributed Systems and Data Mining, minor Degree in Innovation & Entrepreneurship
- First Year: **Computer Science and Engineering**, *Eindhoven Technical University, Eindhoven, the Netherlands*, Final Grade "cum Laude"
- Second Year: **Information and Communication Technology**, *KTH Royal Institute of Technology, Stockholm, Sweden, Final Grade A*

Bachelor's Degree in Computer Science, *University of Trento, Trento, Italy,* 09/2013 - 07/2016
- *GPA 95%, Final Grade 110/110 "cum Laude"*

Awards

- **Merck Research and Innovation Cup 2023** - Placed 2nd with a business case for target prioritization in drug discovery through AI. Accepted as top 1% applicant; Pitched results to Merck executives.
- **IFI – International Research Scholarship** awarded by the DAAD to support a research stay abroad at the RIKEN Research Institute located in Tokyo, Japan.
- **Best Business Plan** award at the European Institute of Innovation and Technology (EIT) Digital Summer School, proposing a congestion-control system for smart cities based on dynamic pricing of roads and parking spots, enabled by real-time analytics and forecasting of traffic patterns.
- **Best Pitch & Business Case** at the Siemens AI@Industry Hackathon, proposing an AI-powered solution to save 20% of the energy utilized by manufacturing production lines. A team in Siemens was assigned to actually develop this product.
- **Winner of the Siemens Tech for Sustainability hackathon**, proposing an AI-powered system to dynamically allocate battery capacity, and projected to reduce industrial electricity costs by 30%.

Publications

- Drost et al., 2024, *Predicting T cell receptor functionality against mutant epitopes*, **Cell Genomics**
- Dorigatti et al. 2023, *Frequentist uncertainty quantification in semi-structured neural networks*, **AISTATS**
- Boniolo et al. 2021, *Artificial Intelligence in Early Drug Development enabling Precision Medicine*, **Expert Opinion on Drug Discovery**
- Fritz et al. 2021, *Combining Graph Neural Networks and Spatio-temporal Disease Models to Predict COVID-19 Cases in Germany*, **Nature Scientific Reports**
- Dorigatti et al. 2020, *Joint epitope selection and spacer design for string-of-beads vaccines*, **ECCB** (best poster award)

(Exhaustive list on Google Scholar)