Emilio DORIGATTI
Data Science, Machine Learning, Statistics, Bioinformatics

in linkedin.com/in/edorigatti   github.com/e-dorigatti
@ emilio.dorigatti@gmail.com

AT A GLANCE

My Vision	Create a fair, sustainable, and thriving society by transforming healthcare through my expertise in Data Science.
AI/ML Skills	Deep learning, Semi-supervised learning, Positive-unlabeled learning, Semi-structured regression, Computer vision, NLP, Uncertainty quantification, Clustering, Representation learning, Interpretable and explainable machine learning, Statistical modeling, Discrete optimization, Visualization, …
Technologies	Python, R, SQL, PyTorch, TensorFlow, TensorBoard, MLflow, Keras, Scikit-Learn, Pandas, Numpy, PyMC3, Spacy, Hugging Face, NLTK, Jupyter, Git, Jenkins, Docker, Spark, AWS, SLURM, Bash, Make, SnakeMake, PostgreSQL, ElasticSearch, Redis, Java, Spring, JavaScript, TypeScript, React, …
Soft Skills	Agile project management, Time management, Leadership, Supervision, Mentorship, Stakeholder management, Communication, Collaboration, Teamwork, Innovation, Entrepreneurship

Throughout my PhD (graduating in Summer 2023) **I worked with international experts from multiple research fields**, including physics, biology, bioinformatics, statistics, mathematics, and computer science, **leading the development of novel data analysis methods**, testing domain-specific hypotheses in order to generate new insights, and **publishing 13 articles** in a variety of domain-specific journals and conferences.

I also participated in the organization and teaching of five University courses and seminars offered to Master's students in Data Science and Statistics, **supervising 18 Master's students** during four group projects, three Master's theses, one Bachelor's thesis, and three seminars. All of this, together with many presentations and my tech blog has taught me how to **communicate complex ideas in simple terms** as well as to understand and **collaborate with people who have widely different skills and capabilities**.

EDUCATION

2019 - 2023	Ph.D. in *Statistics* focused on **Deep Learning**, Uncertainty Quantification, Discrete Optimization. Ludwig Maximilian Universität, Munich, Germany.
2016 - 2018	Master's Degree in **Data Science** with a specialization in *Distributed Systems and Data Mining for Really Big Data*, and a minor Degree in **Innovation & Entrepreneurship**, European Institute of Innovation and Technology (EIT) Digital Master School, GPA 95%
	> 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
2013 - 2016	Bachelor's Degree in **Computer Science**, University of Trento, Trento, Italy, GPA 95%, Final Grade 110/110 "cum Laude"
2008 - 2013	High School Diploma for **Informatics Industrial Engineer**, ITT Marconi Rovereto, Rovereto, Italy, Final Grade 100/100

SELECTED PROFESSIONAL EXPERIENCE

July 2023 July 2019	Machine Learning Scientist and Doctoral Researcher, INSTUTUTE OF COMPUTATIONAL BIOLOGY, HELMHOLTZ ZENTRUM, Munich, Germany
	My Ph.D. project was about developing new frameworks to aid *in silico* **design of personalized vaccines for cancer.** My research focused on quantifying uncertainty in deep learning models, semi-structured regression models, as well as self-supervised and positive-unlabeled learning.
	> **Organized** and **conducted** exercise sessions, exams, seminars and group projects;
	> **Supervised** 18 Master's students during thesis projects, group projects and seminars;
	> **Interdisciplinary collaboration** with other researchers, both as a **team lead** and as a contributor;
	> **Communicated** complex ideas to professionals and neophytes alike, in writing and speech.
	> Additional courses: time management and **agile project management**, good scientific practice, **presentation skills**, scientific writing

June 2018 **February 2018**	**Master Thesis Student, RISE SICS, Stockholm, Sweden** I used machine learning to predict the relationship between the energy that comes off the earth surface and the consequent change in temperature and wind speed with altitude. This is important to make climate simulations more precise, and improve our understanding of Earth's climate. › Found a bug in PySpark and submitted a patch for it; › Benchmarked and interpreted several machine learning algorithms (gradient boosted trees, k-nearest neighbors, neural networks) › Applied and extended my knowledge of big data processing systems (**pySpark**).
June 2017 **March 2017**	**Data Scientist, BRIGHTCAPE B.V., Eindhoven, the Netherlands** As a working student, I consulted a major Dutch bank and helped them develop a system to predict the next actions of their customers to improve staff allocation. › Acted as technical advisor supporting the CTO in **meetings with customers**; › Used predictive process mining for **customer journey analytics**; › Used time series analysis for **demand forecasting** and **capacity planning**.
June 2016 **February 2016**	**Machine Learning Engineer, WIKIDATA, Trento, Italy** For my Bachelor's thesis I improved and extended a text mining framework I developed during the **Google Summer of Code** 2015, resulting in the addition of 4.5M new claims about 12M entities coming from more than 50 different web sources to Wikidata. › Used **web scraping** to obtain textual documents with the required information; › Used **named entity recognition** and **machine learning** to parse the text and extract the new facts;
June 2016 **April 2014**	**Data Engineer, SPAZIODATI SRL, Trento, Italy** I worked on the data acquisition and processing systems underlying the company flagship product. › Created decoupled and extensible APIs through a microservices architecture (**REST APIs**); › Offloaded heavy data processing jobs to a cloud infrastructure (**Spark** on **AWS EC2**); › Gained experience in developing big data systems and working in **agile teams** following **Scrum**.

✎ SELECTED PUBLICATIONS

› Rodemann MJ, Goschenhofer J, Dorigatti E, Nagler T, Augustin T, (2023), *Approximate Bayes Optimal Pseudo Label Selection*, Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence (**UAI**), PMLR

› Dorigatti E, Schubert B, Bischl B, Rügamer D, (2023) *Frequentist uncertainty quantification in semi-structured neural networks*, Proceedings of the 26th International conference on Artificial Intelligence and Statistics (**AISTATS**), PMLR: Volume 206

› Boniolo F*, Dorigatti E*, Ohnmacht A J*, Saur D, Schubert B, Menden M P, (2021) *Artificial Intelligence in Early Drug Development enabling Precision Medicine.* **Expert Opinion on Drug Discovery**, 16:9, 991-1007

› Fritz C*, Dorigatti E*, Rügamer D (2021) *Combining Graph Neural Networks and Spatio-temporal Disease Models to Predict COVID-19 Cases in Germany.* **Nature Scientific Reports** 12 (1), 1-18

› Dorigatti E, Schubert B (2020) *Joint epitope selection and spacer design for string-of-beads vaccines.* **Bioinformatics** 36, i643–i650. European Conference on Computational Biology (**ECCB**)

* Share first/last authorship – Exhaustive list on Google Scholar

◆ AWARDS AND SCHOLARSHIPS

› **Merck Research and Innovation Cup 2023** participant, proposing a business plan to use AI to improve the efficiency and transparency of clinical trials.

› **IFI – International Research Scholarship** awarded by the DAAD to support a research stay abroad at the RIKEN 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% or more.

› **Best Poster** award at the International Conference on Research in Computational Molecular Biology (**RECOMB**) for an epitope-based cancer vaccine design framework, later published in Oxford Bioinformatics.

◉ LANGUAGES

Italian	● ● ● ● ●	English	● ● ● ● ○	German , Chinese	● ○ ○ ○ ○