

Emilio DORIGATTI

Bioinformatics, Data Science, Machine Learning, Statistics

in linkedin.com/in/edorigatti
@ emilio.dorigatti@mail.com

github.com/e-dorigatti
e-dorigatti.github.io



AT A GLANCE

My Vision	Create a fair, sustainable, and thriving society through my expertise in Data Science.
My Values	Reasonable, Reliable, Trustworthy, Competent, Autonomous, Open, Straightforward, Free.
Technologies	Python, R, SQL, PyTorch, TensorFlow, Keras, Scikit-Learn, Pandas, PyMC3, NLTK, OpenCV, Linux, Docker, Spark, AWS EC2/S3, SLURM, PostgreSQL, ElasticSearch, Redis, ...

Throughout my PhD I **worked with experts from multiple research fields**, including physics, biology, bioinformatics, statistics, mathematics, and computer science, **leading the development of novel data analysis methods**, and testing domain-specific hypotheses in order to **generate new insights**.

I also participated in the organization of five university courses and seminars offered throughout the years, **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 personal 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**.

SELECTED PUBLICATIONS

- > 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
- > Ziegler I, Ma B, Nie E, Bischl B, Rügamer D, Schubert B, Dorigatti E, (2022) *What cleaves? Is proteasomal cleavage prediction reaching a ceiling?*, **NeurIPS LMRL** (Learning Meaningful Representations of Life)
- > Dorigatti E, Bischl B, Schubert B, (2022) *Improved proteasomal cleavage prediction with positive-unlabeled learning*, Machine Learning for Health (**ML4H**)
- > Boniolo F*, Dorigatti E*, Ohnmacht A J*, Saur D, Schubert B, Menden M P, (2021) *Artificial Intelligence in Early Drug Development enabling Precision Oncology*. **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
- > Rezaei M, Dorigatti E, Rügamer D, Bischl B, (2022), *Joint Debaised Representation Learning and Imbalanced Data Clustering*, IEEE International Conference on Data Mining (**ICDM**) Deep Learning and Clustering (DLC) Workshop
- > 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**)
- > Dorigatti E, Schubert B, (2020) *Graph-theoretical formulation of the generalized epitope-based vaccine design problem* (RD Kouyos, Ed.). **PLOS Computational Biology** 16, e1008237

* Share first/last authorship

SELECTED PROFESSIONAL EXPERIENCE

Now
July 2019



Ph.D. Student, DEPT. OF STATISTICS, LUDWIG MAXIMILIAN UNIVERSITÄT, Munich, Germany

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

Now
July 2019



Guest Scientist, INSTITUTE 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.

- > Developed vaccine design frameworks based on **mixed-integer linear programming**;
- > Used **deep learning** to improve the identification of neo-epitopes (proteasomal cleavage and MHC binding);
- > Performed **single cell analyses** to quantify the response to vaccines.

<p>April 2019 October 2018</p> 	<p>Digital Consultant, NETLIGHT CONSULTING GMBH, Munich, Germany</p> <p>I consulted a Munich-based financial firm and helped them expand and improve their transaction processing systems.</p> <ul style="list-style-type: none"> › Learned to navigate large, mature code-bases and quickly add value in an agile environment; › Obtained the Payment Card Industry Data Security Standard (PCI DSS) Developer certification; › Acted as a Quality Assurance engineer ensuring reliability in day-to-day operations of the systems.
<p>June 2018 February 2018</p> 	<p>Master Thesis Student, RISE SICS, Stockholm, Sweden</p> <p>I used machine learning to predict, informally, 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.</p> <ul style="list-style-type: none"> › Found a bug in PySpark and submitted a patch for it; › Benchmarked several machine learning algorithms. › Applied and extended my knowledge of big data processing systems (pySpark).
<p>June 2017 March 2017</p> 	<p>Business Analytics Consultant, BRIGHTCAPE B.V., Eindhoven, the Netherlands</p> <p>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.</p> <ul style="list-style-type: none"> › 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.
<p>June 2016 February 2016</p> 	<p>Machine Learning Engineer, WIKIDATA, Trento, Italy</p> <p>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.</p> <ul style="list-style-type: none"> › 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;
<p>June 2016 April 2014</p> 	<p>Data Engineer, SPAZIODATI SRL, Trento, Italy</p> <p>I worked on the data acquisition and processing systems underlying the company flagship product.</p> <ul style="list-style-type: none"> › Created decoupled and extensible APIs through a microservices architecture. › 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.

SELECTED AWARDS

- › **Best Poster** award at the International Conference on Research in Computational Molecular Biology (**RECOMB**) for an epitope-based vaccine design framework, later published in Oxford Bioinformatics.
- › **Winner** of the **Siemens Tech for Sustainability** hackathon, proposing a system of dynamically-allocated batteries to save electricity costs.
- › **Best Pitch & Business Case** at the **Siemens AI@Industry Hackathon**; our solution was projected to save 20% of the energy utilized and 5% of the costs in a manufacturing production line. A team in Siemens was assigned to continue its development and make it a reality.
- › **Best Business Plan** award at the **European Institute of Innovation and Technology (EIT) Digital Summer School**: our team developed a business plan for a congestion-control system in a smart city by using dynamic pricing of parking spots based on real-time analytics and forecasting of traffic patterns.

EDUCATION

2019 - Now	Ph.D. in <i>Statistics</i> 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 <i>Distributed Systems and Data Mining for Really Big Data</i> , and a minor Degree in Innovation & Entrepreneurship , European Institute of Innovation and Technology (EIT) Digital Master School, GPA 95% <ul style="list-style-type: none"> › First Year: <i>Computer Science and Engineering</i>, Eindhoven Technical University, Eindhoven, the Netherlands, Final Grade "cum Laude" › Second Year: <i>Information and Communication Technology</i>, 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