Emilio DORIGATTI
Data Scientist and Artificial Intelligence PhD

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

Professional with 9 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 data science applications to drug discovery, keen to continuously learn and share knowledge as demonstrated by teaching and supervisory experience at a major German University.

▤ AT A GLANCE

AI/ML Skills	Deep learning, Geometric deep learning, Graph neural networks, Semi-supervised learning, Representation learning, Computer vision (CV), Image Segmentation, 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
Soft Skills	Agile project management, Time management, Leadership, Supervision, Mentorship, Stakeholder management, Communication, Collaboration, Teamwork, Innovation, Entrepreneurship

💼 SELECTED PROFESSIONAL EXPERIENCE

July 2023 — July 2019	**Machine Learning Scientist, HELMHOLTZ ZENTRUM, Munich, Germany** 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.
June 2018 — February 2018	**Data Scientist, RISE SICS, Stockholm, Sweden** Improved our understanding of Earth's climate by making simulations more precise; › Developed a big data processing platform leveraging PySpark; › Implemented and benchmarked machine learning algorithms (SVM, Random Forest, XGBoost, GAMs, deep neural networks); › Found a bug in PySpark and submitted a patch for it as a pull request on GitHub
June 2017 — March 2017	**Data Scientist, BRIGHTCAPE B.V., Eindhoven, the Netherlands** 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.
June 2016 — February 2016	**Machine Learning Engineer, WIKIDATA, Trento, Italy** 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.
June 2016 — April 2014	**Data Engineer, SPAZIODATI SRL, Trento, Italy** 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

2019 - 2023   **Ph.D. in Statistics**, Ludwig Maximilian Universität, Munich, Germany.
- Developed optimization frameworks to design vaccines for cancer;
- Studied methodological challenges for artificial intelligence in precision medicine and drug discovery;
- Proposed to unify deep neural networks with statistical regression models to analyze clinical data;
- Additional courses in Project Management and Time Management.

2016 - 2018   **Double Master's Degree in Data Science**, European Institute of Innovation and Technology
- 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.

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

◆ AWARDS AND SCHOLARSHIPS

- **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 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.

✎ 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,* Drost F,* Straub A,*, Hilgendorf P, Bischl B, Busch D, Schober K, Schubert B, (2023) *Predicting T Cell Activation for Mutational Epitopes*, (under review at Nature Immunology)
- Ziegler I, Ma B, Bischl B, Dorigatti E*, Schubert, B*, (2023) *Proteasomal cleavage prediction: state-of-the-art and future directions*, Submitted to **Briefings in Bioinformatics**
- 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