Roben Bhatti

robenbhatti@gmail.com linkedin.com/roben-bhatti github.com/bhroben
+39 331 211 6804 Brescia, Italy Italian citizenship

Experience

German Aerospace Center (DLR) Oct 2024 - Current
Data Scientist Intern *Bremen, Germany*

- Developed and implemented Bayesian models to quantify uncertainty in aerodynamic coefficients for reusable launch vehicles, leading to more robust estimation and performance analysis.
- Built a **CI/CD** pipeline using **Gitlab** and **Docker**, automating the deployment process and ensuring reliable releases.
- Collaborated within an **Agile team** to deliver data-driven results, culminating to the successful completion of my Master Thesis and a **co-authored** conference paper (IAC 2025) detailing the novel application (*in preparation*).
- Addressed computational bottlenecks in model training by implementing Sparse Gaussian Processes, resulting in a **30% decrease in model training time** and enhancing the feasibility of large-scale uncertainty analysis.

Education

University of Padua Oct 2022 - Jul 2025
Master of Science in Physics "Physics of Data" (GPA: 4.00 / 4.00) *Padua, Italy*

- **Relevant Coursework:** Mathematical and Numerical Methods, Deep Learning and Neural Networks, Advanced Statistics for Physics Analysis, Information Theory and Inference, General Relativity.

University of Padua Oct 2019 - Oct 2022
Bachelor of Science in Astronomy (GPA: 3.45 / 4.00) *Padua, Italy*

- **Relevant Coursework:** Advanced Calculus, Statistics, Analytical Mechanics, Quantum Physics, Special Relativity.

Projects

Streaming Particle Physics Data and Online Data Analysis with Spark and Kafka

- Engineered a distributed streaming pipeline capable of processing over **10 GB** of data per day, simulating a real-time detector stream with **Kafka** and **Spark** for live analysis.
- Deployed the environment on a cloud cluster using **AWS S3** for storage, Kafka for message queuing, and Spark for distributed processing.
- Developed a real-time interactive dashboard with Bokeh for **Data Visualization**.

Chess Position Recognition using Transformers (DETR) in PyTorch

- Developed a computer vision pipeline to digitize physical chess games, achieving **87.5%** accuracy in identifying pieces and their positions from a single image.
- Fine-tuned a DETR (DEtection TRansformer) model, leveraging **transfer learning** to significantly reduce training time and improve performance on custom-built dataset.
- Developed a robust **post-processing** module to automatically convert the model's raw detection output into standard notation (FEN), making the board state immediately compatible with chess engines and analysis tools.

End-to-End Machine Learning Pipeline for Profiling Insurance Customers (Ongoing)

- Implemented a medallion data architecture to ensure data quality and scalability for a dataset of **20K** insurance policies.
- Developed a two-stage model to forecast claim amounts: first classifying claim probability, then using a Random Forest regressor to predict the financial amount for high-probability customers.

Technical Skills

Languages: Python, R, SQL
Technologies: Docker, Git, CI/CD, Kafka, Spark, Pytorch, SciPy, NumPy, Pandas, Scikit-learn, Tableau.

Languages

English: C1, **Italian**: Native