Statement of Purpose for Applying to the PhD Program in Computer and Control Engineering

Luca Cotti

June 2024

The rapid advancement of Machine Learning is unlocking unprecedented opportunities to extract insights from vast amounts of data, critical for addressing societal challenges and improving quality of life. My motivation to pursue a PhD is to gain in-depth knowledge of Machine Learning and Data Science, and contribute to developing innovative and real-world technologies that positively impact society. I believe that the research skills I will acquire in designing experiments, analyzing data, and drawing meaningful conclusions are crucial to achieve my goal of pursuing a successful and career in Data Science, as either a researcher or industry professional. Doing so while contributing to the scientific community and society at large is a strong stimulation that I would not find in industry alone.

My motivation stems from my master's thesis, "Study and Development of a Digital Twin for the Simulation of User-Created Automations in a Green Smart Home". In this work I used state-of-the-art data mining techniques to extract operational information of appliances from raw energy data, to then develop a web application to track energy consumption, simulate automation scenarios of appliances, and offer energy-saving suggestions. This project, besides introducing me to the impact of data-driven solutions on energy sustainability and quality of life, also sparked my passion for research, as it resulted in a conference paper for the International Conference on Advanced Visual Interfaces 2024 and a journal article in Future Internet.

My goals are backed by my studies in Computer Science and Engineering at the University of Brescia, where I took courses in Machine Learning and Data Mining. My passion for these fields led me to participate in several Kaggle competitions, applying my knowledge to practical problems. I also completed courses in Operational Research and Optimization Algorithms, culminating in a project on tuning the Kernel Search heuristic for the Multiple Knapsack Problem. Additionally, I gained experience in Cybersecurity through a research project on the implications of Hybrid Work on information security and privacy in organizations. I balanced all of my coursework with the seminars, laboratories, and travels abroad of the Collegio di Merito Luigi Lucchini, where I resided during my studies. Engaging with a community of brilliant minds tought me effective communication, teamwork, and time management—skills that will be invaluable in my doctoral studies.

While moving from my hometown of Brescia to Turin is a significant step, the resources and opportunities at the Politecnico di Torino would be invaluable for my research and professional development. The Smart-Data@PoliTO research center is particularly appealing to me, offering the chance to collaborate with leading researchers in Data Science and Machine Learning. The innovation hub that is Turin itself is also a draw, with its vibrant tech community and numerous opportunities for networking and collaboration.

One area of research I am eager to explore during my PhD is the application of Natural Language Processing techniques to Cybersecurity, specifically to support security analysts against cyber threats. My interest in this topic stems from the potential of Large Language Models to understand structured and unstructured data, which could be harnessed to detect and mitigate AI-powered, persistent, and ubiquitous cyber threats. Valid contributions could include adapting existing log-based models for zero-shot classification of new attack types or expanding log analysis to encompass network traffic, system calls, and other data sources.

Another intriguing research topic is the optimization of transportation systems. This area interests me due to its potential to reduce traffic congestion, improve urban planning, and promote sustainable mobility, significantly impacting energy sustainability, quality of life, and environmental conservation. Contributions could include applying Data Science and Big Data techniques to predict traffic congestion, optimize routes, or study the efficiency of various vehicle types over different terrains and conditions using data from in-car black boxes.