Fairness in AI (automated) Decision Making

Frauke Kreuter[1], Christoph Kern[2]

[1] Department of Statistics, LMU Munich

[2] University of Mannheim

30.06.2021, 16.00 (c.t.)

Online via Zoom
(Meeting-ID: 913-2473-4411; Password: StatsCol21)

The advent of powerful prediction algorithms led to increased automation of high-stake decisions regarding the allocation of scarce resources such as government spending and welfare support. This automation bears the risk of perpetuating unwanted discrimination against vulnerable and historically disadvantaged groups. Research on algorithmic discrimination in computer science and other disciplines developed a plethora of fairness metrics to detect and correct discriminatory algorithms. However, moving theory into practice is full of challenges. For one, data science decisions when implementing prediction models can have large consequences. At a higher level, looking into the philosophical discourse on distributive justice, fairness metrics have their own limitations and can have problematic implications.

This presentation will take a German use case (predicting jobseekers' risk of becoming long-term unemployed) to evaluate commonly used prediction methods on performance, fairness metrics, and vulnerabilities to data analysis decisions. We highlight that different classification policies have very different fairness implications and call for rigorous auditing processes before such models are put to practice. We argue that by cleanly distinguishing between prediction tasks and decision tasks, research on fair machine learning could take better advantage of the rich literature on distributive justice.