NJIT | INSTITUTE FOR DATA SCIENCE

Data Science Seminar Series

Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and use Interpretable Models Instead

Cynthia Rudin, Ph.D.

Professor
Director, Prediction Analysis Lab
Duke University

Date:	Wednesday, March 10th, 2021
Time:	4:00 PM – 5:00 PM EDT
Location:	Zoom Virtual Room
Web Link:	https://njit-institute-for-data-science.eventbrite.com

With widespread use of machine learning, there have been serious societal consequences from using black box models for high-stakes decisions, including flawed bail and parole decisions in criminal justice. Explanations for black box models are not reliable, and can be misleading. If we use interpretable machine learning models, they come with their own explanations, which are faithful to what the model actually computes. I will give several reasons why we should use interpretable models, the most compelling of which is that for high stakes decisions, interpretable models do not seem to lose accuracy over black boxes - in fact, the opposite is true, where when we understand what the models are doing, we can troubleshoot them to ultimately gain accuracy. I will be discussing work from the following papers:

- Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and use Interpretable Models Instead, Nature Machine Intelligence, 2019.
- The Age of Secrecy and Unfairness in Recidivism Prediction. Harvard Data Science Review, 2020.
- Learning Certifiably Optimal Rule Lists for Categorical Data. Journal of Machine Learning Research, 2018.
- Concept Whitening for Interpretable Image Recognition. Nature Machine Intelligence, 2020.

Cynthia Rudin is a professor of computer science, electrical and computer engineering, and statistical science at Duke University, and directs the Prediction Analysis Lab, whose main focus is in interpretable machine learning. She is also an associate director of the Statistical and Applied Mathematical Sciences Institute (SAMSI). Previously, Prof. Rudin held positions at MIT, Columbia, and NYU. She is a three-time winner of the INFORMS Innovative Applications in Analytics Award, was named as one of the "Top 40 Under 40" by Poets and Quants in 2015, and was named by Businessinsider.com as one of the 12 most impressive professors at MIT in 2015. She is a fellow of the American Statistical Association and a fellow of the Institute of Mathematical Statistics.

Some of her (collaborative) projects are: (1) she has developed practical code for optimal decision trees and sparse scoring systems, used for creating models for high stakes decisions. Some of these models are used to manage treatment and monitoring for patients in intensive care units of hospitals. (2) She led the first major effort to maintain a power distribution network with machine learning (in NYC). (3) She developed algorithms for crime series detection, which allow police detectives to find patterns of housebreaks. Her code was developed with detectives in Cambridge MA, and later adopted by the NYPD. (4) She solved several well-known previously open theoretical problems about the convergence of AdaBoost and related boosting methods. (5) She is a co-lead of the Almost-Matching-Exactly lab, which develops matching methods for use in interpretable causal inference.