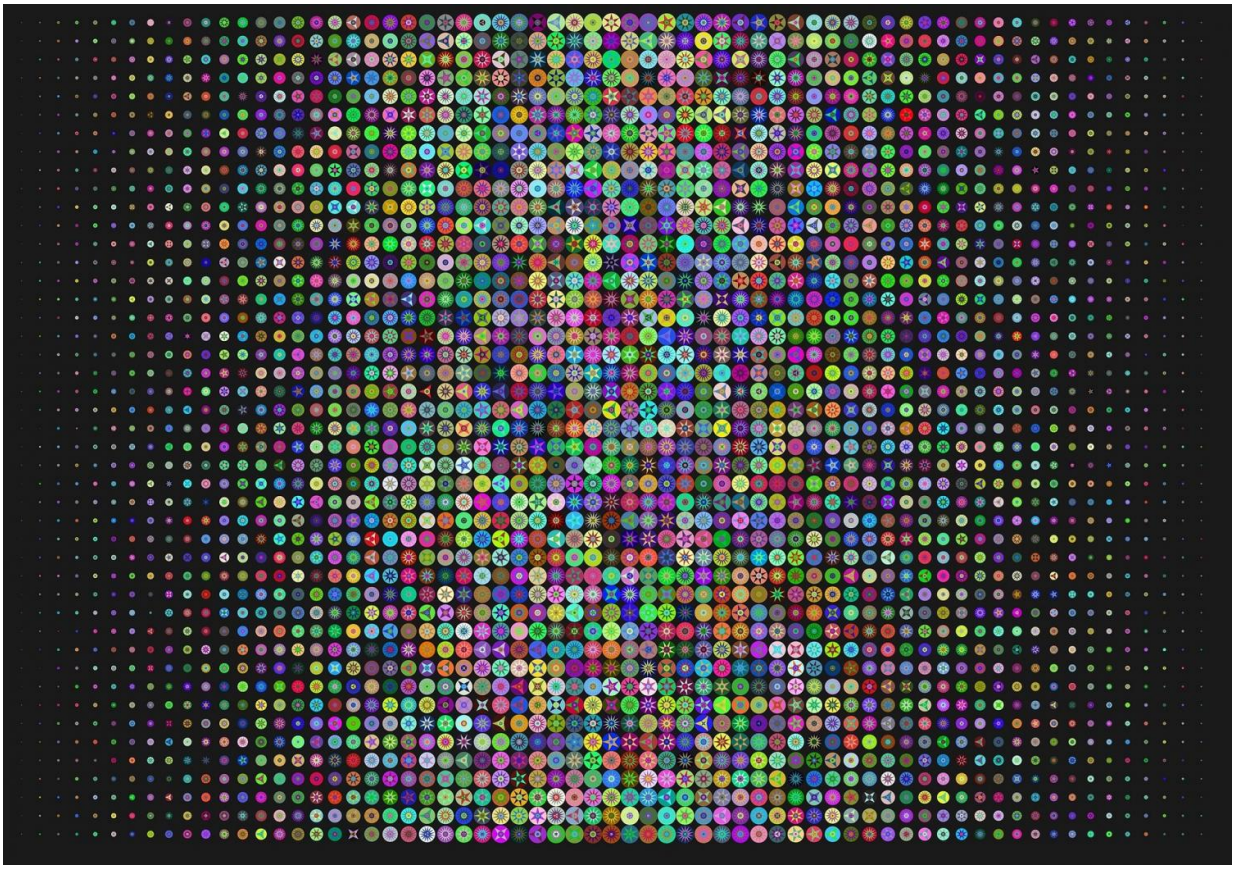
Removing human bias from predictive modeling

October 30 2019



Credit: CC0 Public Domain

Predictive modeling is supposed to be neutral, a way to help remove personal prejudices from decision-making. But the algorithms are packed with the same biases that are built into the real-world data used

to create them. Wharton statistics professor James Johndrow has developed a method to remove those biases.

His latest research, co-authored with his wife, statistician Kristian Lum, "An Algorithm for Removing Sensitive Information: Application to Race-independent Recidivism Prediction," focuses on removing information on race in data that predicts recidivism, but the method can be applied beyond the criminal justice system.

"In criminal justice, there is a lot of use of algorithms for things like who will need to post bail to get out of jail pre-trial versus who will just be let out on their own recognizance, for example. At the heart of this is this idea of risk assessment and trying to see who is most likely, for example, to show up to their court dates," says Johndrow.

"The potential problems with this are just that these algorithms are trained on data that is found in the real world. The algorithms and their predictions can bake in all of this human stuff that is going on, so there has been a lot more attention lately to making sure that certain groups aren't discriminated against by these algorithms."

More information: James E. Johndrow et al. An algorithm for removing sensitive information: Application to race-independent recidivism prediction, *The Annals of Applied Statistics* (2019). DOI: 10.1214/18-AOAS1201

Provided by University of Pennsylvania

Citation: Removing human bias from predictive modeling (2019, October 30) retrieved 28 April 2024 from https://phys.org/news/2019-10-human-bias.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.