

A proposal for a bias and fairness framework that works in practice

Abstract

With the advent of more advanced modelling techniques a need for better explainability emerged and ultimately the desire to assess whether a model is making biased predictions. While currently several tools exist to measure biases in data or models, they do not provide a framework on how to build on any insights gained.

Here, we looking to propose a bias and fairness assessment framework that is separate from any existing tool so that it can lay the foundation for an enterprise-wide bias and fairness management approach that includes the remit of modelling and also extends beyond.

In particular, the framework we present is going to cover the following aspects:

- A clear definition of bias and fairness, their distinction, and what this distinction means for a bank's decision-making process
- Several statistically founded measures to assess bias and guidance on which conclusions that can be drawn from each measure.
- Proposed risk appetite metrics and thresholds to enable clear target setting for fair decision making driven by legal, regulatory, or internal requirements.
- An approach to define a "fairness budget" to quantify an institution's investment into more socially fair credit decisioning as part of their ESG agenda.
- Proposed analyses and visualisation method to assess existing biases before and during model development stages, as well as assessing the effectiveness of active portfolio management decisions to provide fairer credit decisions

Authors & Affiliations

Mr Jakob Kisiala¹

¹True North Partners, Edinburgh, United Kingdom