

Enhancing a Recidivism Prediction Tool with Machine Learning: Effectiveness and Algorithmic Fairness

Authors : Marzieh Karimihaghighi, Carlos Castillo

Abstract : This work studies how Machine Learning (ML) may be used to increase the effectiveness of a criminal recidivism risk assessment tool, RisCanvi. The two key dimensions of this analysis are predictive accuracy and algorithmic fairness. ML-based prediction models obtained in this study are more accurate at predicting criminal recidivism than the manually-created formula used in RisCanvi, achieving an AUC of 0.76 and 0.73 in predicting violent and general recidivism respectively. However, the improvements are small, and it is noticed that algorithmic discrimination can easily be introduced between groups such as national vs foreigner, or young vs old. It is described how effectiveness and algorithmic fairness objectives can be balanced, applying a method in which a single error disparity in terms of generalized false positive rate is minimized, while calibration is maintained across groups. Obtained results show that this bias mitigation procedure can substantially reduce generalized false positive rate disparities across multiple groups. Based on these results, it is proposed that ML-based criminal recidivism risk prediction should not be introduced without applying algorithmic bias mitigation procedures.

Keywords : algorithmic fairness, criminal risk assessment, equalized odds, recidivism

Conference Title : ICAIL 2021 : International Conference on Artificial Intelligence and Law

Conference Location : Toronto, Canada

Conference Dates : June 15-16, 2021