

Title: Survey on Machine Learning Biases and Mitigation Techniques

Author Names: Sunzida Siddique , Mohd Ariful Haque , Roy George , Kishor Datta Gupta , Debashis Gupta and MdJobair Hossain Faruk

Abstract: Machine learning (ML) has become increasingly prevalent in various domains. However, ML algorithms sometimes give unfair outcomes and discrimination against certain groups. Thereby, bias occurs when our results produce a decision that is systematically incorrect. At various phases of the ML pipeline, such as data collection, pre-processing, model selection, and evaluation, these biases appear. Bias reduction methods for ML have been suggested using a variety of techniques. By changing the data or the model itself, adding more fairness constraints, or both, these methods try to lessen bias. The best technique relies on the particular context and application because each technique has advantages and disadvantages. Therefore, in this paper, we present a comprehensive survey of bias mitigation techniques in machine learning (ML) with a focus on in-depth exploration of methods, including adversarial training. We examine the diverse types of bias that can afflict ML systems, elucidate current research trends, and address future challenges. Our discussion encompasses a detailed analysis of pre-processing, in-processing, and post-processing methods, including their respective pros and cons. Moreover, we go beyond qualitative assessments by quantifying the strategies for bias reduction and providing empirical evidence and performance metrics. This paper serves as an invaluable resource for researchers, practitioners, and policymakers seeking to navigate the intricate landscape of bias in ML, offering both a profound understanding of the issue and actionable insights for responsible and effective bias mitigation.

Keywords: machine learning; bias; mitigation techniques; fairness constraints; pre-processing; in-processing; post-processing

<https://doi.org/10.3390/digital4010001>

<https://www.mdpi.com/2673-6470/4/1/1>