Name: Apurba Afiat

Student ID: 23341096

Course Code: CSE424

Section: 01

Review: Predicting Fairness of ML Software Configurations

URL : Predicting Fairness of ML Software Configurations | Proceedings of the 20th International Conference on Predictive Models and Data Analytics in Software Engineering

1 Summary of the paper

1.1 Motivation/purpose/aims/hypothesis: The paper aims to explore the relationship between machine learning (ML) hyperparameters and fairness, particularly in the context of ensuring equitable outcomes in ML applications. The hypothesis is that by understanding this relationship, it is possible to predict the fairness of hyperparameter configurations, thereby improving the training process and reducing bias.

1.2 Contribution: The authors contribute a data-driven framework that utilizes regression methods to predict the fairness of ML hyperparameters. They provide empirical evidence demonstrating that this prediction is feasible for fixed datasets with specific protected attributes.

1.3 Methodology: The study employs a series of experiments using 30 different benchmarks across five ML algorithms and six fairness-sensitive tasks. The performance of various ML prediction algorithms (DNN, SVR, TR, and XGB) is evaluated based on metrics such as Root Mean Square Error (RMSE) and the coefficient of determination (R^2).

1.4 Conclusion: The findings indicate that it is possible to predict the fairness of ML hyperparameters effectively, which can lead to more efficient training processes by avoiding biased configurations. The authors highlight the importance of hyperparameters in influencing the accuracy, robustness, and fairness of data-driven software.

2 Critiques or limitations

2.1 1st Critique/Limitation: The approach relies heavily on the quality and diversity of the input dataset, which is known to be a significant source of discrimination. The authors acknowledge that their method cannot eliminate all fairness issues, as it focuses primarily on the training process.

2.2 2nd Critique/Limitation: The study uses only two group fairness metrics (AOD and EOD), which may not capture the full spectrum of fairness issues. This limitation could lead to scenarios where a model is deemed fair while still adversely affecting vulnerable groups.

2.3 3rd Critique/Limitation: The reliance on heuristics in the Parfait-ML tool for generating hyperparameter configurations may not guarantee the discovery of optimal or interesting hyperparameters within the given time constraints, potentially limiting the effectiveness of the proposed framework.

3 Synthesis

3.1 1st potential/idea of a new/follow-up/extension paper: A follow-up study could investigate the integration of additional fairness metrics and their impact on the prediction of hyperparameter fairness. This could involve developing a more comprehensive framework that considers various dimensions of fairness beyond AOD and EOD.

3.2 2nd potential/idea of a new/follow-up/extension paper: Another potential extension could explore the application of the proposed framework in different domains, such as healthcare or criminal justice, where fairness is critical. This research could assess how the framework performs with diverse datasets and protected attributes, providing insights into its generalizability and adaptability.