Peer Review of Debang Ou's Research Paper on Recidivism Forecasting

Andrew Resnick
University of Connecticut

November 14, 2024

1 Review

To my understanding, the goal of the research paper is to generate a predictive model to better identify individuals at higher risk for recidivism, and to enhance decision-making surrounding this process in the criminal justice system. The author explores Logistic Regression, LASSO, and Random Forest (RF) to fulfill the addressed gap in advanced models for predicting inmates at higher risk for recidivism. The author concludes that RF is the most fair and accurate model in this case, showing a high accuracy and relatively low false positive and false negative rates. Lastly, the author rightfully addresses the need to evaluate fairness in these advanced models as to not reinforce biases regarding race, economic status, and other factors. The major comments on this paper are as follows:

- 1. Address fairness in the models: In the initial questions proposed, one was to, "How accurate and fair are Logistic Regression, LASSO, and Random Forest models in predicting recidivism?". However, the assessment of fairness within the models used seemed to not be discussed later in the paper. Although this is partly addressed in Section 4.3 for future research, it may be beneficial to remove or address fairness directly within the evaluation of your models.

- 2. Discussion of assumptions: Although Section 1 is expansive, it would help the reader to discuss the assumptions of each model and to evaluate whether the assumptions of each model are fulfilled within this

context. Since this research paper is not only targeted toward statisticians, but those who can utilize these models in decision making, adding a section about assumptions of the models used may add much needed context to readers not familiar with each model.

- 3. Brief discussion of hypotheses: The research questions are clearly articulated, but the paper doesn't clearly identify hypotheses for these models. Including a hypothesis and then evaluating the results within the context of the hypothesis can help non-statisticians with some context surrounding the performance of the models.

The author did a very good job proofreading the paper, as I did not catch any small mistakes or typos within the paper. This was a very interesting paper with high impact!