Nikita McClure Midterm

10/25/2024

Pursuing Open-Source Development of Predictive Algorithms: The Case of Criminal Sentencing Algorithms argues that open-source penalized regression algorithms are in some ways superior to "black-box" algorithms and should be used, especially in the case of criminal sentencing. Specifically, the paper argues that specific open-source algorithms(ridge regression, LASSO regression, and elastic net regression), are superior to the commonly used black-box algorithm COMPAS for criminal sentencing. They argue that this is the case due to several factors: increased accuracy, decreased racial bias, improved affordability, potential for collaboration, and the ability for the public to analyse and test the algorithm. Throughout this essay, the methods that the researchers utilized to prove the aforementioned features will be elaborated upon and their importance will be discussed. When a person is convicted of a crime, the judge must decide upon a suitable punishment for their crime. Judges must consider a variety of factors in this decision, one of which is, if the person is released, whether or not they will commit another crime, or "recidivate". This is a very difficult thing for a human to predict; the opportunity to utilize algorithms to assist in this analysis can beis invaluable. The severity of punishment has a large impact on the course of the convict's future. It impacts all areas of their life, including employment, family planning, and their future lifestyle. Due to the incredible importance this decision has on a person, it should be as appropriate to the individual as possible. Having the opportunity to utilize an algorithm that can help the judge predict the likelihood of recidivism can take a huge burden off the judge and might make it more fair for the individual.

However, this is only the case if the algorithm itself is a fair and valid tool for the judge to utilize. An algorithm's fairness can be improved by being an open-source algorithm, rather than a black-box due to its transparency and collaboration potential. As open-source algorithms are public, anyone who desires can view, test, and assist in contributing to it. This means that the algorithm has much potential to be continuously and significantly improved upon because many minds from various disciplines can contribute to it–likely resulting in many more iterations. Also, due to its transparency, defendants have the opportunity to see why they are being sentenced how they are. Arguments have been made,though none have yet been upheld, that the exclusion of this reasoning due to the use of black-box algorithms violates the fair procedure component of due process. Open-source algorithms eliminate this potential issue.
Additionally, and possibly most significantly, this study demonstrates that each of these specific open-source algorithms can already be shown to be both more accurate and less biased than COMPAS. Findings have shown that COMPAS has a lower accuracy when predicting for Black individuals than for white individuals. Additionally, while inaccurate predictions for White people tend to predict lower recidivism rates, Black people are more often incorrectly predicted to re-offend. This means that the COMPAS algorithm is more likely to promote longer sentences and less chance of parole for Black individuals than for White. The lower cost of the open source application should make it more readily available. The more the same decision tree is utilized, the more consistent and thus "fair" the outcomes should be. COMPASand other closed algorithms are privately owned and utilize proprietary information. Because of this, their use and improvements can be very expensive. While maintenance of an open-source algorithm does accrue expenses, its use overall is very affordable, especially in comparison. This can be valuable as it provides the opportunity for under-resourced court systems to benefit from its use, where the more costly option would be more difficult to access. The necessary training for its use is also less time-intensive and costly than possible decision-making training for humans. This means that the algorithms can be used widely and normalize how punishments are decided. Because every case would adhere to the same parameters, this could reduce the risk of inconsistent and unreliable application, making it more fair overall.

COMPAS alone is more widely used than all the open source algorithms. The prevailing use of COMPAS can be attributed to a few reasons: it is established, both legally through policy and historically, by being the first

such product used in many courts, there is lack of familiarity and trust in the open-source alternatives, and there are data privacy concerns with open-source algorithms. Closed-source algorithms are done completely in house, the personal data used to train and test the algorithm is available only within the company. On the contrary, the training and testing data for open-source algorithms are publically available, though they are anonymized. There is concern that the data used in the algorithm could be de-anonymized, exposing personal information of hundreds of individuals. Despite the advances in data science that have allowed very innovative and thorough ways to anonymize data, decoding is always a risk. The specific open-source algorithms being discussed only use: race, sex, age, juvenile felony/misdemeanor/"other" crime classification counts, and counts of all non-juvenile misdemeanors and felonies, all of which was obtained from public records. This means that in the specific case we are discussing, there is no risk of exposing or violating a person's privacy. In similar cases, where information used is more sensitive, the value of the privacy of an individual must be weighed against the potential good that the algorithm, and it being open-source, can do. In a case like this, when the algorithm aids in overall equality of such an extreme decision, I believe the good to the individuals and the system outweighs the bad done to the people losing privacy. John Mill's theory of utilitarianism dictates that the open-source algorithm does overall more good than bad, so it must be implemented.

Waggoner and Macmillen ran a series of tests to support their hypothesis that the open-source algorithms ridge regression, LASSO regression, and elastic net regression are more accurate overall and less biased than the COMPAS algorithm. The study used 7214 real criminal records from Broward County Florida. These records included information such as gender, age, home zip code, prior offenses, and information on if the individual reoffended within 2 years. Waggoner and Macmillen replicated how COMPAS would predict recidivism for each offender using the COMPAS prediction feature. They compared the replicated predictions to the observed outcomes – if recidivism actually occurred. This was used to analyze the accuracy, the baseline accuracy was found to be 65.4%. This was used as the baseline to compare each open-source algorithm. Additionally, by addressing the inaccuracies in prediction trends for various racial groups, bias was assessed.

Each open-source model was fit to a penalized regression algorithm (shown in Figure 1).

Notably, each algorithm has a penalty parameter, λ to adjust or shrink coefficient estimates to improve model performance. For each algorithm, to find λ, a K-fold cross-validation, with K of 10, was conducted. A partition of 80% training data and 20% testing data was used. The training subset was fit to a base model, each testing subset was used to predict the outcomes with various λ values. The λ value that was on standard deviation greater than the minimum λ was selected for each algorithm.

Ridge Regression (equation 1) uses λ to tune theℓ_2-norm penalty. This penalty shrinks the coefficient estimates to very small but does not drop them completely.

Lasso regression (equation 2) uses ℓ_1-norm penalty, tuned by λ. ℓ_1, in contrast to ℓ_2, penalizes the absolute value of the coefficient estimates. This results in redundant features within the specification being dropped.

Elastic net regression (equation 3) uses both ℓ_1 (Lasso) and ℓ_2 (Ridge) penalties, in this case blending the two using a tuning parameter, α. Selecting for this parameter is done following the same 10-fold cross-validation methods used for λ, this time predicting for various λ and α. The λ and α selected were those that jointly minimized the mean squared error".

The mean predictive accuracy across 1,000 iterations was then measured for each model. Each of the open-source algorithms performed slightly better than COMPAS. While COMPAS had an accuracy of 65.4% (according to this study, others range from 61%-66%), LASSO was 67.2%, Ridge was 67.1%, and Elastic Net was 67.4%.

Confusion matrices were also generated for each model in order to observe true positive, true negative, false positive, and false negative rates. Stronger performance was indicated by higher true positive and true negative results.

ROC curves were plotted for each model to visualize the trade-off between true and false positive rates. For all of the open-source models, the area under the curve was between 0.73-0.74, meaning each algorithm had reasonable predictive accuracy.

Lastly, each model was examined to investigate which features contributed the most to predicting recidivism rates. While ridge regression retained all features, LASSO did not include certain features like charge degree and juvenile misdemeanor counts. Additionally, elastic net " "combined the L1 and L2 penalties to balance feature selection to retain only the most predictive features while minimizing model complexity.

To analyse the validity of this study, I focused on the LASSO algorithm. I did so because all three algorithms were very close in accuracy, but LASSO specifically had the lowest predicted false positives, which are considered to be the worst outcome. I replicated the algorithm in Rstudio using data from the same dataset as the original paper, to confirm the declared overall accuracy of the results. I obtained an accuracy of 67.77%, higher than the 67.2% declared by the authors. However, this was likely due to the fact that I could not differentiate testing versus training data for this replication. I used the original dataset as there is very little publicly available data with information regarding juvenile criminal defenses. Overall this accuracy, and the process of the algorithm, minimizing potential redundant or less relevant factors. An additional aspect I provided was analyzing the accuracy and the distribution of true and false positives and negatives of White offenders versus Black offenders. While this paper discusses the racial disparities of Black offenders receiving false positives at a higher rate from the COMPAS prediction. It does not assess if the algorithms they present have the same issue. After analyzing the accuracy, I generated confusion matrices of true/false-negative/positive predictions for the overall algorithm as well as one specific to White offenders and one specific to Black offenders. I found that while the overall accuracy was potentially sufficiently good, the race disparities were consistent throughout the LASSO algorithm as well, though in a different way. The accuracy of overall prediction for both race groups was very close, and in fact slightly more accurate for Black offenders (67.5% White, 67.6% Black). White offenders were more often accurately predicted to not reoffend than Black offenders, 52.89% and 36.36%, respectively. However, the accuracy is still very close because Black offenders were more accurately predicted to offend, 31.28% and 14.63% respectively. On the contrary however, Black offenders were less often predicted to reoffend when they did not than White offenders (20.16% versus 24.74%). While the accuracy for reoffending was much higher for Black individuals, the false positives, arguably the most important factor was not. The LASSO algorithm is not perfect and may in fact be working too hard to fight racial bias against Black individuals which may result in the pendulum swinging too far in the other direction. However, the steps being made is progress, and the fact that I was able to access the algorithm to verify the results myself is large benefit of open-source algorithms In conclusion, Pursuing Open-Source Development of Predictive Algorithms: The Case of Criminal Sentencing Algorithms demonstrates that open-source penalized regression models, specifically ridge, LASSO, and elastic net regression, have many advantages over proprietary "black-box" algorithms like COMPAS. This is especially the case in high-stakes scenarios such as criminal sentencing. Through improved transparency, lower costs, increased accuracy, and reduced bias, open-source algorithms allow for more fair and reliable decision-making. By comparing each algorithm and its predictions on real-life data the study confirms that open-source methods are often as good as or better than "black box" algorithms in accuracy and ethical transparency. These findings support adopting open-source algorithms for predictive purposes especially in contexts where unbiased results are essential such as in our justice system.

Appendix:

$\hat{\beta}_{\lambda}^{ridge} = argmin\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} B_j X_{ij})^2 + \lambda \sum_{i=1}^{p} \beta_j^2\}$

$\hat{\beta}_{\lambda}^{LASSO} = argmin\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} B_j X_{ij})^2 + \lambda \sum_{i=1}^{p} |\beta_j|\}$

$\hat{\beta}_{\lambda}^{EN} = argmin\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} B_j X_{ij})^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2\}$