

Taking Out the Guesswork: An Analytical Approach to Police Traffic Stop
Contraband Searches

Tong Zhou

Anderson Summer Science Program/Data Analytics Department
Research Advisor Dr. Anthony Joseph Bonifonte

The purpose of this research is to develop a model to prevent unnecessary vehicle contraband searches for the police, better identify targets that likely contain contraband, and decrease social disparities among different demographic groups. To build this model, we used a data set provided by Stanford University (*E. Pierson, C. Simoiu, J. Overgoor, S. Corbett-Davies, V. Ramachandran, C. Phillips, S. Goel. (2017)*). The data contains detailed information about police traffic stops and searches for contraband (such as drugs). We separated the data into training sets (2006-2013) and test sets (2014-2015). Taking reasons to stop, historical contraband find percentage in the county, time of the day, and day of week as predictors, our logistic regression model shows the police are more likely to find contraband in the evenings, on weekends, and in counties with high history of contraband found. In addition, even though we never include gender as a variable when we built the model, the data shows the police searched males almost six times more than females, despite females having higher contraband found percentage. In our test set, our logistic regression model will conduct searches equally, in proportion to the observed rates of contraband found. Our predictive model generated a test AUC score of 0.691 and accuracy of 0.832. The contraband find percentage in the top 10% of predicted risk

is 28.6%, compared with 19.3% in the overall data. Furthermore, among 30% of the highest predicted contraband probability, our model identifies 64.8% of the total contraband found. In conclusion, our model can give police a reliable baseline estimate of whether a vehicle contains contraband or not.