

Predicting the offender's age

Stephan Stahlschmidt^{1,2,*}

1. Humboldt-Universität zu Berlin, School of Business and Economics, Chair of Statistics, Spandauer Straße 1, 10178 Berlin, Germany

2. Ruprecht-Karls-Universität Heidelberg, Centre for Social Investment, Adenauerplatz 1, 69115 Heidelberg, Germany

*Contact author: stahlschmidt@wiwi.hu-berlin.de

Keywords: Offender Profiling, Prediction Power, Statistical Learning

Predicting the age of an unknown offender is notoriously difficult for police profilers. These experts rely solely on traces found on the crime scene and combine this information to deduce the exact events which happened on the crime scene. Only afterwards they try to characterise the offender based upon the unassured knowledge drawn from the crime scene. These statements on the offender are often based on assumptions and lack certitude. However, an assured prediction on, for example, the offender's age would prove very useful to narrow the group of potential suspects.

We therefore compare the performance of several prediction techniques on a data set of sex-related homicides. The applied techniques include linear regression, k-Nearest-Neighbour, regression trees, Random Forest and Support Vector Machine. We evaluate each approach by its prediction power concerning the offender's age and account for the specific requirements of forensic data, namely the restricted access to information via the crime scene and uncertainty of available information. Our results show, if and how the police's investigation could be enhanced by implementing one of these well-known prediction techniques.

References

- Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33, 1-22.
- Hornik, K. , Buchta, C. and Zeileis, A. (2009). Open-Source Machine Learning: R Meets Weka. *Computational Statistics*, 24, 225-232.
- Karatzoglou, A., Smola, A., Hornik, K. and Zeileis, A. (2004). kernlab - An S4 Package for Kernel Methods. *Journal of Statistical Software*, 11, 1-20.
- Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2, 18-22.