

## Factors affecting crimes in Sri Lanka

**C.B. Wickrama<sup>1</sup> and L.S. Nawarathna<sup>1,2\*</sup>**

<sup>1</sup>Postgraduate Institute of Science, University of Peradeniya, Sri Lanka, <sup>2</sup>Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka

\*lakshikas@pdn.ac.lk

Crimes have been disturbing threats to all the Sri Lankans all over the country. Crimes prevention should be achieved by altering the policies by government agencies. Finding key variables associated with crimes are very vital for policy makers. In this study, our main goal was to forecast of homicides, rapes and counterfeiting currency from 2013 to 2020 using auto regressive conditional Poisson (ACP) and auto regressive integrated moving average (ARIMA) models. All the predictions are made assuming that the prevailing conditions in the country affecting crime rates remain unchanged during the period. Moreover, multiple linear regression analysis and lasso regression were used to identify the key variables associated with crimes. Profiling of districts as safe or unsafe was performed based on overall Sri Lanka's total crime rate which is to compare with individual district's crime rates. Data were collected from Department of Police and Department of Census and Population. There are 14 safe and 11 unsafe districts in Sri Lanka. Besides it is found that total migrant population and percentage of urban population is positively correlated with total crime according to the partial correlation analysis. In time series analysis, ACP (1, 1) models were selected for homicides, rapes and counterfeiting currency. In Multiple linear regression analysis, total migrant population, population density, percentage of Sinhala nationalities are significant variables for total crimes. Forecasts were made for Kurunegala and Anuradhapura districts and both forecasts were within 95% prediction intervals for total crimes. Total migrant population, Gini index, percentage of urban population, percentage of Sinhala nationalities are significant variables for homicides in multiple linear regression model. Forecasts were made for Colombo and Gampaha districts and both predictions were within 95% prediction interval for homicides. In lasso regression, total migrant population, population density, percentage of Sinhala nationalities are significant variables for total crimes. Furthermore, in the analysis of homicides, total migrant population, Gini index, population density and percentage of Sinhala nationalities are significant. Random K-nearest neighbor algorithm classified districts as safe and unsafe without use of the actual crimes and predicted with 84% of accuracy.