Application of R in Crime Data Analysis

Anna Dyga*, Monika Sławińska**

I. Kielce University of Technology
*Contact author: anna_dygaa@o2.pl
**Contact author: monikaslawinska@o2.pl

Keywords: Crime, Forecasting, Decision Tree,

The use of $R$ in the context of crime data analysis certainly contrasts with its most popular application – in finance, economics, or planning and prediction which are largely connected to trade or services. It is common knowledge that in many countries crime rate has been on the rise over the last few years and it is a complicated issue that governments struggle to solve. The aim of our presentation is to introduce the application of $R$ program in the police's statistics and show its numerous advantages. $R$ is an open source program and the use of it is fairly simple and does not necessarily require advanced IT abilities. This makes it an ideal means to deal with the enormous amount of data gathered by the police, to simplify its processing and interpretation. For the sake of this presentation, we have analyzed two data sets with the use of $R$. The first set contains data about crimes committed in 2010 in particular countries. The second set is concerned with data about crimes in all the regions of Poland. To analyze the data, we have used easily-comprehensible histograms which show the crime rate in every country. Besides, to depict which crimes are prevalent and which are committed less often, we have used mosaic plots available in rattle package. However, the main goal of our presentation is to specify the most dangerous areas and to this end we have employed decision trees with the help of packages tree and rpart. We have used this particular tool because it is lucid and greatly simplifies the reasoning. As countries included in the first data set are at various stages of development (developed, developing and least-developed), it has been possible to try to point out a relationship between the economic advancement and citizens' tendency to commit crimes. What is more, we have discovered whether types of crimes committed in those least-developed countries are in any way similar to or different from those which dominate in countries with high-functioning economy. For greater clarity, presentation of our findings and interpretation of them is followed by maps depicting crime rates in particular areas both in Poland and around the world. Moreover, the analysis of the second data set has enabled us to establish the extent of crime in Poland in comparison to other countries, and to learn whether Poland is among the countries of higher crime rate or it is relatively safe to visit the country. It is necessary to point out that $R$ program can be further employed in the analysis of data concerning smaller units, namely city districts – a beneficial method especially in case of metropolises with diversified national and cultural makeup, such as London or New York City where multiculturalism may be a source of tension between people and create conditions conducive to crime rise. We are convinced that application of $R$ would make crime forecasting easier and increase the ability to deal with crime on local and global scale.

References