Data mining with Rattle

Milena Nowek¹*, Justyna Jarmuda¹**

1. Kielce University of Technology, al. Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Poland
*Contact author: milena.nowek@gmail.com
**Contact author: juska88@wp.pl

Keywords: Rattle package, data mining, useful and clear information, calcium-silicate bricks

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining software, such as package Rattle, is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationship identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases [1]. Data mining combines concept, tools, and algorithms from machine learning and statistics for the analysis of very large datasets, so as to gain insights, understanding and actionable knowledge [2].

R is ideally suited to the many challenging tasks associated with data mining. Rattle (the R Analytical Tool To Learn Easily) is a graphical data mining application written in and providing a pathway into R. It has been developed specifically to ease the transition from basic data mining to sophisticated data analyses using a powerful statistical language. Rattle brings together a multitude of R packages that are essential for the data miner but often not easy for the novice to use. An understanding of R is not required in order to get started with Rattle – this will gradually grow as we add sophistication to our data mining [2], and this is what will be shown in our poster presentation – the way of exploring the analysis tools offered by Rattle, from basic to more advanced. We will try to present the most relevant statistical studies associated with marketing research and engineering datasets, both the essential and more complex. Our analyses are focused on modified calcium-silicate bricks. The main purpose is to present how a big number of information from huge databases can be combined, reduced and shown in the form of graphs and diagrams, and in what specific case it might be helpful. Simply saying – how to obtain the most desirable, useful and clearly expressed information from a large table of results of conducted research, using package Rattle.

References
