

Symptom Checklist-90-Revised Rating Scale: A Data Mining Approach

Evgenia Gkintoni¹, Magdalena Nikiel², Sampson Fytros² Constantinos Halkiopoulos^{2,3}, Gerasimos Antzoulatos^{2,3}

¹*Department of Psychology, University of Crete, Greece*

²*Department of Business Administration, Technological Educational Institute of Western Greece, Greece*

³*Department of Mathematics, University of Patras Artificial Intelligence Research Center, University of Patras, Greece*

In this paper were applied Machine Learning and Data Mining methods to capture the psychological age group of students 18-26 years. For recording, tracing and evaluation of the psychological condition, was used the standardized scale Symptom Checklist-90 (SCL-90), which examines a wide range of psychological problems and symptoms of psychopathology.

The methodology adopted, in first phase consists of electronic questionnaires, which were created and posted through the website <http://www.cicos.gr>. Subsequently data were collected and preprocessed from the questionnaires and then introduced into the Weka (Waikato Environment for Knowledge Analysis) Machine Learning Platform for analysis and extraction of useful knowledge. More specifically, through using classification algorithms (ID3, C4.5) there was a production of prospectively decision trees. Decision trees are a powerful way in order to represent and facilitate statements analysis (psychological) principally, comprising successive decisions and variable results in a designated period.

Furthermore, clustering technique (*K-Means* algorithm), was applied, which is a well-known knowledge discovery process of extracting previously unknown knowledge, actionable information from very large scientific and commercial databases. The *k-means* is a very popular algorithm and one of the best for implementing the clustering process. Also, the parameters of the algorithm were set, depending on the application cases, and also the results were correlated with the birth-place and the place of present residence, educational background of both the respondents and first-degree relatives, professional occupation of parents and other parameters, in order to evaluate and assess the significance of exported rules / conclusions. In addition, the respondents were classified into clusters based on 9 clinical signs (subscales) of the scale SCL-90.

The results indicate among others, that the use of Data Mining methods is an important tool to export and receive the conclusions and decisions especially in the field of psychological assessment and in neuroscience.