

CLASSIFICATION OF A LIFE INSURANCE DATABASE WITH SPSS

Brunela Karamani¹, Esteriana Haskasa², Jozef Bushati³

¹Polytechnic University of Tirana, Faculty of Information Technology, Tirana Albania
E-mail: b.karamani@gmail.com

² University of “Aleksandër Xhuvani”, Department of Mathematics, Elbasan ,Albania
E-mail: esteriana.haskasa@gmail.com

³ University of “Luigj Gurakuqi”, Shkodra
E-mail: jozefbushati@gmail.com

Abstract

Recently, especially over the last two decades there is an exponential increase of the amount of information and data stored in electronic format. The large amount of electronic information is saved in databases, which sizes are increasing from day to day. The development and application of data mining algorithms requires the use of powerful software tools. As the number of available tools continues to grow, the choice of the most suitable tool becomes increasingly difficult. Implicitly this situation poses some challenges. These challenges include the introduction of new techniques that complement traditional methods of statistical analysis, and which allow the detection and interpretation of knowledge in very large databases. This paper presents an application of a classification tree building process, one of the most common data mining techniques. For that we used CHAID, Exhaustive CHAID, CART and QUEST algorithms, which are implemented in specialized software in data mining - SPSS. The life insurance dataset is used to study the information and bring out the hidden knowledge from it. All algorithms first applied on training dataset and created the decision tree, pruning method used for reducing the complexity then rule set are derived from decision tree. Same rules then applied on evaluation data set. Comparing the results of all algorithms and recommended the appropriate product to the new customer those having similar characteristics. Then, there are presented the summary table related with the created models and decision trees diagrams. The performance of each classifier model is evaluated by using statistical measures like accuracy, specificity and sensitivity and gain chart.

Keywords: *database, information, techniques, algorithm, statistical analysis, life insurance.*