

A Study on the Application of Data Mining Techniques for Classification and Clustering of Medical Data

Christos Koukouvinos

Department of Mathematics, National Technical University of Athens, Zografou, Athens, 15773, Greece. Email: ckoukouv@math.ntua.gr

Efthalia Massou

Department of Mathematics, National Technical University of Athens, Zografou, Athens, 15773, Greece. Email: efthalia.massou@gmail.com

Kalliopi Mylona

Department of Mathematics, National Technical University of Athens, Zografou, Athens, 15773, Greece. Email: kmylona@math.ntua.gr

Abstract

Data mining is the science of extracting nontrivial, previously unsuspected and finally comprehensible information from large databases and applying it for decisions making. This new discipline plays an essential role in exploring and interpreting massive medical data sets. This paper is concerned with the application of data mining techniques to the analysis of the trauma annual data in Greece for the year 2005. The data set consists of 6334 records, 25 variables and a binary response variable (death or not). In our study, different data mining techniques are implemented and decision trees, classification rules and clusters are produced. The results of C&RT, CHAID, C5.0 and QUEST are evaluated not only before but also after the implementation of feature selection methods in the examined data set. For clustering, EM and K-means algorithms are used to identify valuable clusters of records.

Keywords: Data mining, medical data sets, decision trees, classification rules, clusters, feature selection.

2010 Mathematics Subject Classification: 62-07, 62P10.