

ABSTRACT
USAGE OF CLUSTER ALGORITHMS IN HEALTH STUDIES:
AN APPLICATION

Özge PASİN

Master of Science Thesis, Biostatistics and Medical Informatics Department

Supervisor: Prof. Dr. Handan ANKARALI

SEPTEMBER 2015, 240 Pages

With clustering methods variable and individuals which have similar characteristics may be collected in a group. Although clustering methods have many applications, there are limited studies in health researches in our country. While the purpose of this study is to introduce different clustering algorithms and show how and which cases should be correctly used. At the same time, different clustering algorithms results which can be applied on a real data set were compared. According to the evaluations, for two different data sets the kappa coefficients were statistically significant and its degree are intermediate. In terms of both data sets the most convenient and fastest algorithm is *Farthest* clustering algorithm. The results obtained by *Make Density Based* and *EM* algorithms gave the most accurate decisions in terms of the distribution of the groups among Framingham risk groups crosstables. As a result, with taking into account the criterion of clinical information it is thought that the examination of clustering of risk factors of the disease, will be played an important role for introduction of accurate disease diagnosis. In addition we believe that when considering data distribution and characteristics of data sets clustering algorithms can be used as a diagnostic tool for the plannings and diagnosis of diseases in the field of health.

Key Words: Algorithm, Framingham Risk Score, Cluster Analysis, Data Mining