Abstract

In this paper, we scrutinize factors that dole out significantly to augmenting the risk of hepatitis-C virus. The dataset has been taken from the machine learning warehouse of University of California. It contains nineteen features along with a class feature having binary classification. There is a total of 15 binary attributes together with a class attribute and 5 continuous attributes. The dataset contains 155 records. In order to prevail over the missing
values problem, data normalization techniques are applied. First, the dimension of the problem is trimmed down. Next binary logistic regression is applied to classify the cases by using qualitative and quantitative approaches for data reduction. The three stage procedure has produced more than 89% accurate classification. Our proposed approach has a low feature complexity with a good classification rate as it is working by using only 37% of the total fields.

Reference

- Dunham M. H. and Sridhar S., Data Mining: Introductory and Advanced topics, Pearson Education, 2006

**Index Terms**

Computer Science

Biomedical

**Key words**

Binary logistic regression analyses
data mining
hepatitis-C Virus (HCV)

principle component analysis