Abstract

For processing of large amount of data numerous techniques are used. Data Mining is one of the technique that is used most often. To process these data, Data mining combines traditional data analysis with sophisticated algorithms. Medical data mining is an important area of Data Mining and considered as one of the important research field due to its application in healthcare domain. Classification and prediction of medical datasets poses real challenges in Medical Data Mining. To cope with these challenges Logistic Regression (LR) and Artificial Neural Network (ANN) are commonly used. LR enables us to investigate the relationship between a categorical outcome and a set of explanatory variables. LR explains that there can be one or more independent variables that can determine the problem outcome. ANN resembles the human brain and here the information is processed by simple elements called neurons and signals are transmitted between the neurons. Feature subset selection selects subsets of features that are enough to explain the target concept. In this paper feature selection methods like forward selection and backward elimination using mean evaluation are used on the medical datasets. LR and ANN are applied on feature selection methods using Cross Validation Sample (CVS)
and Percentage Split as test options. From the experimental results it is identified that for SPECTF dataset LR using percentage split prediction accuracy of 83.95% is achieved, for Diabetes Dataset LR using percentage split prediction accuracy of 80.46% is achieved, and for Liver Disorder dataset NN using percentage split prediction accuracy of 74.75% is achieved.

References


13. Ankita Dewan and Meghna Sharma, "Prediction of Heart Disease Using a Hybrid
Technique in Data Mining Classification”, 2015, page(s): 704-706.

Index Terms

Computer Science
Artificial Intelligence

Keywords

Cross Validation Sample, Data Mining, Mean Evaluation, Feature Subset Selection, Logistic Regression, Artificial Neural Network, Percentage Split.