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

1. Sunita Soni, Ujma Ansari, Dipesh Sharma and Jyoti Soni, “Predictive Data Mining for
Medical Diagnosis: An Overview of Heart Disease Prediction”, International Journal of
Computer application (0975-8887), vol. 17, no.8, March (2011).
2. Raghavendra B.K., Jay B. Simha, “Performance Evaluation of Logistic Regression and
Neural Network Model with Feature Selection Methods and Sensitivity Analysis on Medical Data
Mining”, International Journal of Advanced Engineering Technology (Vol. II, Issue: I,
3. Raghavendra B.K., S.K. Srivatsa, Raghavendra S, Shivashankar S.K., “Comparison of
Logistic Regression and Neural Network Model with and without hidden Layers”, Universal
4. Qi Cheng, Pramod K. Varshney, and Manoj K. Arora, Logistic Regression for Feature
Selection and Soft Classification of Remote Sensing Data”, Geoscience and Remote Sensing
5. Raghavendra B.K., Jay B. Simha, “Evaluation of Logistic Regression Model with Feature
Selection on Medical Dataset”, International Journal of Computational Intelligence (Vol.1, Issue
2, Dec 2010), pp. 35-42.
6. Qinbao Song, Jingjie Ni and Guangtao Wang, “A Fast Clustering Based Feature Subset
Selection Algorithm for High Dimensional Data”, IEEE Transactions on Knowledge and data
7. Amin S.U., Agarwal, K. and Beg, R.," Genetic neural network based data mining in
prediction of heart disease using risk factors", IEEE Conference on Information &
Communication Technologies (ICT), Page(s):1227– 1231, 2013.
9. Rachata N., Charoenkwan P., Yooyativong T. Chamnongthai K., Lursinsap C. and
Higuchi, K.” Automatic Prediction System of Dengue Haemorrhagic Fever Outbreak Risk by
Using Entropy and Artificial Neural Network”, International Symposium on Communications and
Venugopal, “Classification of Alzheimer’s Disease and Parkinson’s Disease by Using Machine
Learning and Neural Network Method", Second International Conference on Machine Learning
Disease Using Classification Data Mining Techniques”, IEEE International Conference On
Advances In Engineering.
12. R.Robu and C. Hora”, “Medical Data Mining with Extended WEKA“, IEEE International
13. Ankita Dewan and Meghna Sharma, "Prediction of Heart Disease Using a Hybrid


Index Terms

Computer Science
Artificial Intelligence

Keywords

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