Abstract

The major work in data pre-processing is handling missing value imputation in Hepatitis Disease Diagnosis which is one of the primary stages in data mining. Many health datasets are typically imperfect. Just removing the cases from the original datasets can fetch added problems than elucidations. An appropriate technique for missing value imputation can assist to generate high-quality datasets for enhanced scrutinizing in clinical trials. This paper investigates the exploit of a machine learning technique as a missing value imputation process for incomplete Hepatitis data. Mean/mode imputation, ID3 algorithm imputation, decision tree imputation and proposed bootstrap aggregation based imputation are used as missing value imputation and the resultant datasets are classified using KNN. The experiment reveals that classifier performance is enhanced when the Bagging based imputation algorithm is used to foresee missing attribute values.

References
An Ensemble approach on Missing Value Handling in Hepatitis Disease Dataset

3. Information regarding hepatitis C from the staff of Mayo Clinic; available at: http://www.mayoclinic.com/health/hepatitis-c/DS00097
14. Liu Peng, Lei Lei , A Review of Missing Data Treatment Method
15. http://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm

Index Terms

Computer Science Information Sciences

Keywords
data mining, prediction, knn, imputation, missing values, bagging, bootstrap