Abstract

Decision trees and Random Forest are most popular methods of machine learning techniques. C4.5 which is an extension version of ID.3 algorithm and CART are one of these most commonly use algorithms to generate decision trees. Random Forest which constructs a lot of number of trees is one of another useful technique for solving both classification and regression problems. This study compares classification performances of different decision trees (C4.5, CART) and Random Forest which was generated using 50 trees. Data came from OECD countries health expenditures for the year 2011. AUC and ROC curve graph was used for performance comparison. Experimental results show that Random Forest outperformed in classification accuracy [AUC=0.98] in comparison with CART (0.95) and C4.5 (0.90) respectively. Future studies more focus on performance comparisons of different machine learning techniques using several datasets and different hyperparameter optimization techniques.
23. Rotim S.T. Dobsa J. Krakar Z. “Using decision trees for identification of most relevant
indicators for effective ICT Utilization”, Bulgarian Academy of Sciences, Cybernetics and Information Technologies, 2013, 13(1), 83-94.

Index Terms

Computer Science

Algorithms

Keywords

Pattern Recognition, Machine Learning, Decision Trees, Health Expenditures.