World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:12, No:09, 2018

Optimization of Machine Learning Regression Results: An Application on Health Expenditures

Authors : Songul Cinaroglu

Abstract : Machine learning regression methods are recommended as an alternative to classical regression methods in the existence of variables which are difficult to model. Data for health expenditure is typically non-normal and have a heavily skewed distribution. This study aims to compare machine learning regression methods by hyperparameter tuning to predict health expenditure per capita. A multiple regression model was conducted and performance results of Lasso Regression, Random Forest Regression and Support Vector Machine Regression recorded when different hyperparameters are assigned. Lambda (λ) value for Lasso Regression, number of trees for Random Forest Regression, epsilon (ε) value for Support Vector Regression was determined as hyperparameters. Study results performed by using 'k' fold cross validation changed from 5 to 50, indicate the difference between machine learning regression results in terms of R^2, RMSE and MAE values that are statistically significant ($p < 0.001$). Study results reveal that Random Forest Regression (R^2 ˃ 0.7500, RMSE ≤ 0.6000 ve MAE ≤ 0.4000) outperforms other machine learning regression methods. It is highly advisable to use machine learning regression methods for modelling health expenditures.

Keywords : machine learning, lasso regression, random forest regression, support vector regression, hyperparameter tuning, health expenditure
Conference Title : ICBDAKD 2018 : International Conference on Big Data Analytics and Knowledge Discovery
Conference Location : Dublin, Ireland
Conference Dates : September 06-07, 2018