The Best Prediction Data Mining Model for Breast Cancer Probability in Women Residents in Kabul

Authors : Mina Jafari, Kobra Hamraee, Saied Hossein Hosseini

Abstract : The prediction of breast cancer disease is one of the challenges in medicine. In this paper we collected 528 records of women's information who live in Kabul including demographic, life style, diet and pregnancy data. There are many classification algorithm in breast cancer prediction and tried to find the best model with most accurate result and lowest error rate. We evaluated some other common supervised algorithms in data mining to find the best model in prediction of breast cancer disease among afghan women living in Kabul regarding to momography result as target variable. For evaluating these algorithms we used Cross Validation which is an assured method for measuring the performance of models. After comparing error rate and accuracy of three models: Decision Tree, Naive Bays and Rule Induction, Decision Tree with accuracy of 94.06% and error rate of %15 is found the best model to predicting breast cancer disease based on the health care records.
Keywords : decision tree, breast cancer, probability, data mining