

ABSTRACT

Lower Back Pain (LBP) is a very common condition experienced across various demographics. The "Years Lived with Disability" (YLDs) caused by Lower Back Pain increased by 54% from 1990 to 2015, a rise commonly attributed to population growth and aging. The data mining process involves several stages, one of which is preprocessing. This research applies the Kendall feature selection method to choose the most relevant variables based on the strength of their non-parametric ordinal association with the target variable. The objective of this study is to determine the extent of the performance change, in terms of accuracy, on a model that uses the Kendall feature selection technique on the Lower Back Pain dataset. The research uses the Random Forest classification algorithm and employs K-Fold Cross Validation to obtain a suitable value. This study compares a model built without feature selection against one that uses feature selection. The results show that the model without Kendall feature selection achieved an accuracy of 80,71%, while the model using Kendall feature selection achieved an accuracy of 83,27%.

Keyword: *Feature Selection Kendall, Random Forest, Data Mining, Lower back Pain.*