FEATURE SELECTION HYBRID METHOD FOR OPTIMIZATION ALGORITHMS IN THE FIELD OF MEDICAL RECORDS

Yuda Syahidin, Ade Irma Suryani

Manajemen Informasi Kesehatan, Rekam Medis dan Informasi Kesehatan, Piksi Ganesha Politechnic, Bandung, Indonesia

Abstract

Medical Records or Electronic Health Records refers to the collection of patient health information in a digital format. The problem of classifying medical record data involves high-dimensional features. This raises a problem in determining which features have a correlation with the predicted results. Embedded technique uses learning model construction and feature selection simultaneously. The Wrapper technique performs feature evaluation by utilizing machine learning algorithms. The experimental results produce accuracy values for each embedded and wrapper technique. In this research, it is proposed to develop a hybrid technique that aims to find feature significance by applying machine learning techniques to increase the accuracy of predictions for disease classification. The proposed hybrid model combines the results of feature weighting and compares feature performance with several known classification techniques. The test results resulted in an increase in the accuracy value according to the disease classification dataset through the hybrid feature weight evaluation (HFWE) model.

Keywords: Electronic Health Records, Embeded, Wrapper, Hybrid

Introduction

The health of patients is a priority issue and medical experts are constantly trying to implement new technologies and achieve important results. The use of medical data data that enables the analysis of large volumes of medical data that can be used in research areas such as clinical decision support phenotypic information extraction disease inference and personalized healthcare(Shickel dkk, 2017). Predictive analytics is one of the important areas of medical science to provide better services to patients (Panesar, n.d.). In recent years most of the methods used to analyze data from electronic health records which contain a lot of information about patient health have used predictive analytics to extract information using machine learning and statistical techniques. such as the results of clinical trials (Jensen dkk, 2012),(Xiao dkk, 2018),(Dinov, 2018). The performance of prediction algorithms depends on data representation and feature selection (Bengio dkk, 2013),(Polyzotis dkk, 2018). The challenge in health data is to find patterns that generate predictive models to support clinical decisions and much health data remains untapped. Electronic health reporting still lacks the scope and efficiency of medical record data analysis and because there is still unstructured health data there is still a lack of systems for health decision making(Goldstein dkk, 2017), (Weiskopf dkk, 2013), (Latif dkk, 2020).

There are additional approaches to feature selection using filter and banding methods to implement a stepwise search strategy to discriminate feature quality and feature evaluation. All the features in the data sheet are important for establishing a model hypothesis for prediction. The filtering method is a selection method that has an independent machine learning method and is a selection method based on the relationship of variables and a machine learning algorithm. A feature selection study used convolution techniques to reduce the features selected by Yang et al.(Yan & Zhang, 2015), Feature Selection Packing Techniques in Diabetes estimate, Le, et al (Le dkk, 2021), cluster ranking technique in feature selection by Anwar et al (Haq dkk, 2019), and Genetic algorithms used in feature selection to help predict mortality patterns by Ghorbani et al (Ghorbani dkk, 2020).

However, these different feature selection methods due to inappropriate data reduction methods and limitations for feature selection still have challenges with high-dimensional data problems and require more in-depth research and continuous research to combine features. Based on this there is a need for methods to develop feature selection algorithms to facilitate feature selection for medical record data with both structured and unstructured features.