Proceedings of the 11th Annual World Conference of the
Society for Industrial and Systems Engineering,
2022 SISE Virtual Conference
October 6-7, 2022

Predicting Heart Disease and Reducing Survey Time Using Machine Learning Algorithms

S. Rababa, A. Yamin, S. Lu, and A. Obaidat

State University of New York, Binghamton
Binghamton, NY 13902, USA

Corresponding author's Email: srababa1@binghamton.edu

Abstract: Currently, many researchers and analysts are working toward medical diagnosis enhancement for various diseases. Heart disease is one of the common diseases that can be considered a significant cause of mortality worldwide. Early detection of heart disease significantly helps in reducing the risk of heart failure. Consequently, the Centers for Disease Control and Prevention (CDC) conducts a health-related telephone survey yearly from 400,000 participants. However, several concerns arise regarding the reliability of the data in predicting heart disease and whether all of the survey questions are strongly related. This study aims to utilize several machine learning techniques, such as support vector machines and logistic regression, to investigate the accuracy of the CDC's heart disease survey in the United States. Furthermore, we use various feature selection methods to identify the most relevant subset of questions that can be utilized to forecast heart conditions. To reach a robust conclusion, we perform stability analysis by randomly sampling the data 300 times. The experimental results show that the survey data can be useful up to 80% in terms of predicting heart disease, which significantly improves the diagnostic process before bloodwork and tests. In addition, the amount of time spent conducting the survey can be reduced by 77% while maintaining the same level of performance.

Keywords: Classification, Data Sampling, Feature Selection, Heart Disease, Imbalanced Data, Machine Learning

ISBN: 97819384962-3-3