

Lung Cancer Survival Prediction Using Random Forest Based Decision Tree Algorithms

Zakarie Hashi

E-mail: zakarie.a.hashi@wmich.edu

¹Department of Statistics, Western Michigan University, Kalamazoo, MI.

Rabia Almamlook

E-mail: rabiaemhamedm.almamlook@gwmich.edu

¹Department of Industrial Engineering, Western Michigan University, Kalamazoo, MI.

Abstract

lung cancer is the leading cause of cancer death in both men and women in the United States. The main objective of this paper is to analyze the lung cancer data available models to lung cancer survivability prediction model and to develop accurate survival prediction models using Machine Learning. Decision Tree (DT) and Random Forest (RF) algorithms have been applied for constructing a lung cancer survivability prediction model. The Decision Tree and Random Forest classifiers developed in this work predicted the various factors that influence the survival time, would help doctors make more informed decisions about treatment plans and help patients develop more educated decisions about different treatment options. This study has explained the survival rate analysis of patients with advanced lung cancer who did not receive any type of therapeutic modality and to evaluating performance scores daily activities the results of this study have found slight improvement in survival rates. Random Forest Based Decision tree algorithms were found to result in the best prediction performance in terms of accuracy and area under the ROC curve. Overall, these of this show that the algorithms have a promising result for this purpose with the overall survivability prediction model with an accuracy of 85%.

Key words: Random Forest, boosting, bagging, Decision tree, survival analysis, advanced lung cancer