Understanding Lung Cancer Survival Factors Through Advanced Analytics

Opara, Daniel (School: Prospect Hill Academy Charter School High School)

The use of advanced analytics, when applied properly, can offer a greater understanding of its topic of study. For example, advanced analytics in healthcare data can be used to analyze patient trends in certain diseases, discover which practices are most effective to treat illnesses and evaluate health care practitioners. In this specific project, advanced analytics was applied to public domain healthcare data to generate insight on the survival of lung cancer patients. To elaborate, after performing variable analysis to determine meaningful survival predictors and cleaning up the public domain data, the data was used in the Kaplan-Meiers, Nelson-Aalen, and Cox-Proportional Hazard Models to perform survival analysis. Utilizing the Kaplan-Meier Model aided in generating plots whose curve measured the fraction of patients living for a certain amount of time after the start of the clinical trial, thus producing the probability of survival. Comparatively, the Nelson-Aalen Model—also referred to as the Hazard function—provides the instantaneous death rate given the time t. The Cox-Proportional Hazard Model, on the other hand, is a regression model that outputs the association between the survival time of patients and predictor variables that are quantitative, non-categorical predictor variables and for categorical variables. Through the use of the Kaplan-Meier and Nelson-Aalen model, a survival estimate was graphed for the clinical study data used in this project.

Awards Won: