

P8130_final_project

Yifei Yu

2024-12-03

Abstract

Introduction

The data we used for this analysis originates from a breast cancer survival dataset collected from a prospective study. The dataset contains 14 important variables, including patients' age, race, marital status, tumor size, cancer stages (T Stage, N Stage, and 6th Stage), estrogen and progesterone status, and regional node involvement. The dataset also records patients' survival times in months and their final survival status (dead or alive). Our objective is to develop models that predict the risk of death among breast cancer patients using these features. Specifically, we aim to identify which variables significantly impact patients' survival outcomes, determine (potential interactions among the variables, and assess the performance of the models across racial groups). Additionally, we will investigate fairness in the model predictions to ensure equitable accuracy between the majority race group "White" and the minority "Black".

Methods

Results

Conclusion/Discussion

Group members' contribution