

STAT 432 Final Project

Ruining Tao (rtao6), Elsie Wang (erqianw2), Zhan Shi (zhans5)

Dec 10, 2021

Contents

Project description and summary	2
Literature Review	2
Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes	2
Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer	2
Tumor characteristics and patient outcomes are similar between invasive lobular and mixed invasive ductal/lobular breast cancers but differ from pure invasive ductal breast cancers	3
Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome	3
Summary Statistics and Data Preprocessing	3
Data Overview	3
Remove Missing Values	4
Deal with Multicollinearity	4
Continuous Variables	4
Categorical Variables	5
Modeling PR Status	6
Support Vector Machine (SVM)	6
Random Forest	7
Modeling Histological Type	8
Logistic Regression	8
Logistic Regression with Ridge Penalty and Cross Validation	9
K mean Clustering with Principal Components Analysis	10
Classification Tree	11
Variable Selection for All Outcomes	11
Literature	11
Fit Random Forest to Select Variables	11
Generate Fold ID	11
Predict All Four Outcomes	11