Titanic Survival Prediction by Machine Learning

Ningjing Gao

University of oklahoma

May 4, 2020

Abstract

The sinking of Titanic caused thousands of passengers died. The lack of lifeboats and the mistakes in rescue measures are the main reasons for the death of so many passengers in this disaster. The data collected from https://www.kaggle.com/c/titanic/data shows the features like the gender, the age even the ticket class level decides whether they can survive. This paper will base on the passengers feature to build machine learning models and Naive Bayes Classifiers to analysis. This can predict who has more possible to survive. Then use the model to predict whether or not the passenger survived the sinking of the Titanic.

1 introduction

Through the movie Titanic, the ”Titanic” disaster on April 15, 1912 is known all over the world. As in the movie, a small number of lifeboats can only save a small number of passengers. First-class passengers, women and children are given priority. This article will make predictions about which passengers will survive the sinking through machine learning. This algorithm can predict different combinations of survival functions. The data analysis will be completed after the application of the algorithm and the accuracy will be checked. Different algorithms are compared based on accuracy and best performance. It is recommended to use the model for prediction.

2 Research Finding

Depending on the bar chart, we can learn the probability of 3rd class for people not survived is larger than 1st and 2nd class, and the 1st class for survived is higher than others. Hence, the relationship between survived and Pclass is dependent

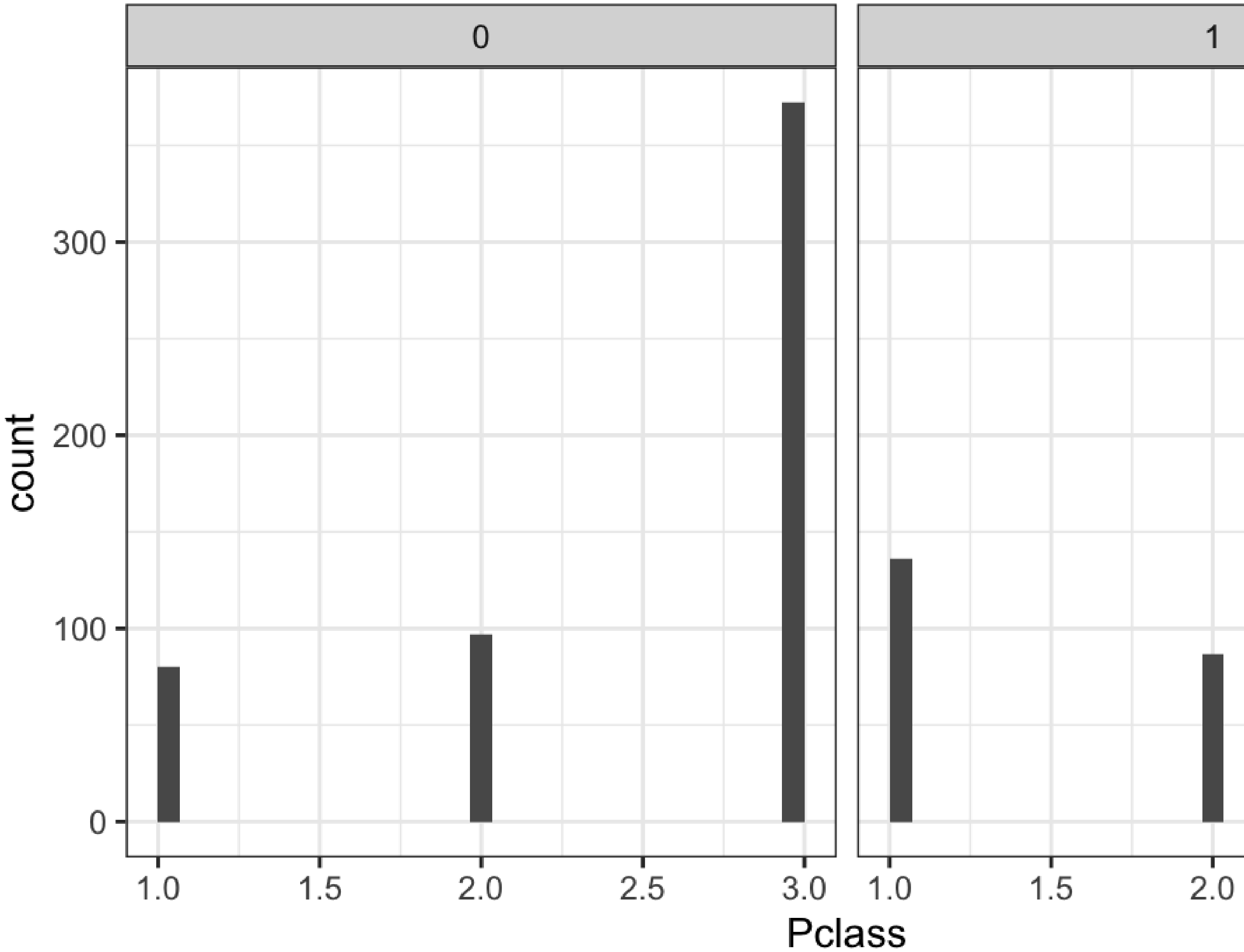

Figure 1: The Survival Probability of The Different Ticket Class

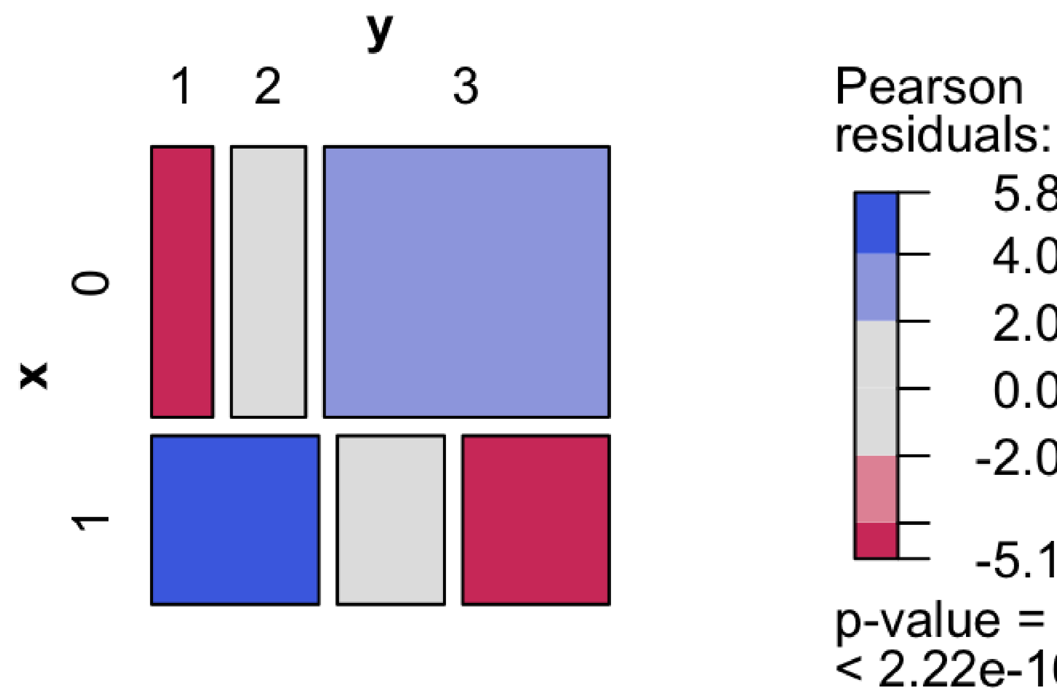

Figure 2: visualize the relationship between survival status and passenger class

References

Table 1: Survival Predictions for Each Combination of Pclass,Sex Values

Panel : Sex = female

	Pclass1	Pclass 2	Pclass3
Survived0	0.003367003	0.006734007	0.080808081
Survived1	0.102132435	0.078563412	0.080808081

Panel : Sex = male

	Pclass1	Pclass 2	Pclass3
Survived0	0.086419753	0.102132435	0.336700337
Survived1	0.050505051	0.019079686	0.0527497191

Panel : Sex = female

	Pclass1	Pclass 2	Pclass3	sum
Survived0	0.003367003	0.006734007	0.080808081	0.090909091
Survived1	0.102132435	0.078563412	0.08080808 1	0.261503928
sum	0.105499439	0.085297419	0.161616162	0.352413019

Panel : Sex = male

	Pclass1	Pclass 2	Pclass3	sum
Survived0	0.086419753	0.102132435	0.336700337	0.525252525
Survived1	0.050505051	0.019079686	0.052749719	0.122334456
sum	0.136924804	0.121212121	0.389450056	0.647586981

Panel : Sex = sum

	Pclass1	Pclass 2	Pclass3	sum
Survived0	0.089786756	0.108866442	0.417508418	0.616161616
Survived1	0.152637486	0.097643098	0.133557800	0.383838384
sum	0.242424242	0.206509540	0.551066218	1.000000000

Notes: The prediction using these predictors, Pr[Survival—Pclass, Sex], is simply a function of Pclass,Sex.

Table 2: Survival Predictions for Each Combination of Pclass

	Pclass1	Pclass 2	Pclass3
Survived0	0.08978676	0.10886644	0.41750842
Survived1	0.15263749	0.09764310	0.13355780

Notes: survived0 is not survived, survived1 is survived.

Table 3: marginal distributions for both survival status and passenger class

	Pclass1	Pclass 2	Pclass3	sum
Survived0	0.08978676	0.10886644	0.41750842	0.61616162
Survived1	0.15263749	0.09764310	0.13355780	0.38383838
sum	0.24242424	0.20650954	0.55106622	1.00000000

Notes: survived0 is not survived, survived1 is survived.