UDACITY

PROJECT

Titanic Survival Exploration

A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Meets Specifications

Hi

First let me congratulate you on your first step towards learning this fascinating area of programming! You've done excellent work on this project and you've clearly understood the material in the tutorials so far. Great job documenting the process that you used to write predictions_3. Being able to articulate your results and how you achieved them is a very important skill in machine learning. Really excellent work so far! Congratulations on passing this project and good luck with the next section of the course!

Cheers!

Answers to Each Question

The `predictions_0` function has been run and the accuracy of the predictions is reported.

Great job!

This result means that if you always guessed that a passenger did not survive, you would be right more often than not (more often than random chance). Another way of thinking about this is that the dataset is unbalanced. Neither of the outcome categories (survived vs. didn't survive) is represented proportionally in your dataset. In the case of two classes of outcomes, proportional representation would be ~50% for each outcome, ~33% if there were three classes etc.

Recognizing this type of situation will become very important when you start splitting your data up into separate subsets for training and testing an algorithm. For instance, you wouldn't want to train a learning algorithm on a subset of the data where 75% of the passengers didn't survive, and then test your results on the remaining subset where ~55% of the passengers survived. You'd want to split the data in such a way that any inherent proportions or biases are preserved across each subset. You'll learn how to deal with this later in the course, but it's always a good first step to use a dummy estimator to get a sense of any biases or trends in your data.

The `predictions_1` function has been correctly implemented. The expected accuracy of the predictions is reported.

Correct!

You've just demonstrated that women were more likely to have survived the Titanic than men. What do you think would be a likely reason for this? Are women generally stronger swimmers than men? Better suited for cold temperatures? Could there be some environmental factor that would have helped women more than men?

These sorts of questions relate to something in machine learning called *domain knowledge*. Domain knowledge refers trends and biases affecting a dataset that aren't inherently part of the dataset themselves. In this case, we know that women were much more likely than men to end up in the lifeboats. Having domain knowledge is frequently the difference between success and failure in your machine learning endeavors!

The `predictions_2` function has been correctly implemented. The expected accuracy of the predictions is reported.

Exactly!

Your result indicates that women and children were significantly more likely to survive than men. According to research that has been done on shipwrecks, this is actually the exception rather than the rule. The Titanic was unusual in that the captain was able to enforce that women and children were given access to the lifeboats before men. However, in most shipwrecks, it's first come, first serve:

http://www.livescience.com/21951-women-children-first-shipwreck-myth.html

The `predictions_3` function has been correctly implemented and obtains a prediction accuracy of at least 80%. The approach to the task has been documented, including features that were explored and intermediate steps taken to complete the function.

Great job improving your classification accuracy! The Pclass variable is also useful for identifying males who were likely to have survived.

A valid scenario where supervised learning can be applied is reported. A clear outcome variable and at least two potential predictor variables are identified as part of the description.

Nice example! We could also use categories of previous purchases that the potential customer was known to have made as a predictor variable.

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review

Student FAQ Reviewer Agreement