Module 5 Homework

2022-02-04

Module 5 Assignment

For this assignment, I'd like you to take a look at the data set `compas-scores-two-years-w-marital-bw-cleaned.csv`, which is linked above. This data set was created from the data provided by ProPublica in their article "Machine Bias: Risk Assessments in Criminal Sentencing" (propublica.org). You can find the original data and some explanation of how they did their analysis linked from that article as well. However, to complete this assignment, you don't necessarily have to read their analysis document. Note that for simplicity's sake, our data set was limited to just "African-American" and "Caucasian" subgroups.

The ProPublica article analyzes the performance of a proprietary risk assessment algorithm from Compas, using data from Broward County, Florida. The data in our data set doesn't necessarily match the data that was available to Compas, but instead represents the data that ProPublica was able to obtain from public records as it worked to assess the results of the Compas algorithm. Remember from our lectures that, according to Pro Publica, the Compas algorithm achieved an overall accuracy of about 59–63% in predicting whether an arrestee would be rearrested within two years. Furthermore, the following percentages describe the true positive rate (sensitivity) and true negative rate (specificity) on arrestees labeled "African-American" and "Caucasian":

Reported Race	Measurement	Percent
African-American	True Negative Rate (Specificity)	55%
African-American	True Positive Rate (Sensitivity)	72%
Caucasian	True Negative Rate (Specificity)	77%
Caucasian	True Positive Rate (Sensitivity)	52%

Your overarching goal in this assignment is to **create a predictive model that at least reaches the overall accuracy of the Compas model (63%) while delivering results that are not biased with respect to African-American and Caucasian arrestees.** In other words, you'd like the sensitivity and specificity for those two groups to be the same. I was able to reach that goal in my code (but perhaps not without using variables that shouldn't necessarily be included in the model), but it's OK if you don't get there. **A secondary goal is to document the efforts you made and the results you obtained in order to better understand the difficulties inherent in creating unbiased machine learning models.**

Specifically, to complete this assignment, you should create and knit an R Markdown document that does the following things:

1. Load the data set and perform any necessary cleaning. While the data set should be relatively "clean" already, you may want to think about which variables are numbers, which should be factors, which should be character, etc.

2. Decide on which variables you will consider for your model. There are likely more variables that you can easily use. Which do you believe might be useful? Write the code to select those variables, and in the text include a short explanation for why you chose the variables you did. (Note: you may decide later to only use some of these variables, but this step is about identifying the largest set of variable's you'll consider.)

3. Create tables or graphs to visualize the values of your variables. Ideally, data visualizations should start to give you an idea of which variables are important and how they might relate to the response variable and each other. At minimum, you should do the following.

    a) Create tables or graphs to understand the distributions of **each individual variable**. For example, are there the same number of African-American and Caucasian arestees in the data set?

    b) Create graphs that look at how each explanatory variable is related to the response variable `two_year_recid`.

    c) Write a paragraph or two that points out any features of your tables or graphs that you believe are important. You don't need to describe every graph, but rather comment on aspects that might affect your analysis later.

4. Create a training and test set from your original data. (You can decide to use cross-validation for model creation and assessment if you like, but it's not required for this assignment.)

5. Using your training set, create at least three predictive models that attempt to predict `two_year_recid` using some or all of your chosen predictor variables.

    a) For each model, include a paragraph or so that explains why you chose that particular model, or those particular variables, and any other decisions you made that might be relevant. Ideally, your explanation might also describe how you hope the next model will improve on the previous model. You might consider overall choice of model (logistic regression, SVM, random forest, etc.) as well as choices made to balance classes or categories.

    b) Write the code to make the model and to assess its overall accuracy and its sensitivity and specificity. You could write your own code, pull the numbers out of the `confusionMatrix` command output, or find other packages if need be. (Note: It should be relatively easy to match the accuracy of the Compas algorithm. We might take a minute to ponder what that means. It may be harder to get unbiased results for the two racial groups.)

6. Finally, summarize the results of your three models in an easy-to-read table, and comment on what you found as a result of this process.

7. Submit both your RMD file and a PDF file of the final knit document.

A few final notes:

- If you're stuck, it is possible to ask for a few hints to get you started.
- It would be good to reflect on whether you were able, with limited data, to do as well as the commercially-available product, in terms of accuracy? Furthermore, is that accuracy good enough for the intended purpose?
- It might be interesting to look at whether your algorithm of choice actually finds the "race" variable very useful.

- I mentioned that I succeeded in fulfilling the goal of the assignment, but that I had to "cheat," in a sense. For me that meant that race was involved in making the final call. In your final comments, discuss whether you think that's "cheating" in this real-world example. Is it OK to use race as a variable if your goal is to create a racially unbiased model?
- Because this is an assignment where you might not "succeed," writing about the process is important.
- To receive full points, an assignment should take some effort to relate each choice of model to specific bias-reduction goals. Perhaps variable weighting was implemented to solve a certain problem. Or perhaps the set of explanatory variables was chosen with a particular goal in mind. Maybe your choice won't achieve the stated goal, but be sure to document why you made the choice. Or, said another way, a submission that could be summarized by "I tried a logistic regression, a random forest and a support vector machine, each with the same variables" would not completely fulfill the objectives for this assignment.