Objective

Explore the explainable classifier outputs and continue thinking about bias, variance, and fairness issues as they pertain to a simplified version of a real-world application, then discuss such issues in a small group.

Clarification

The demographic explanation has a somewhat annoying known issue that seems to have arisen from changes to the pandas library since I originally wrote it. If you run into errors re-running the notebook, you may find that commenting out those two function calls from the display helper will make the problem go away. The error should not harm your ability to complete the assignment, however, because the demographic explanation is global, meaning the previously-canned output is sufficient for you to see what these look like.

Recipe Ingredients

(BEGIN same as lab 7) For reference the code presented in this lab is analogous to that found in the paper presented in the class prior to this labLinks to an external site.. You can find the original posting about this dataset from ProPublica hereLinks to an external site., if you want to learn more about any features. *(END same as lab 7)*

1. Explanation Generation Notebook, this is the canvas file downloadDownload this is the canvas file download
2. One particular test/train split of our original recidivism data (as before, place these in a folder called "data" right next to the pre-existing "sample_data" folder): test_3.csv Download test_3.csv, test_new_3.csv Download test_new_3.csv, train_3.csv Download train_3.csv, and train_new_3.csvDownload train_new_3.csv
3. Form small groups to discuss the "Team" elements of the task list below (Feel free to also discuss anything else that is related to this lab+assignment). It might be easier to retain the groups from lab 5. If you do not have a group, use the discussion board or other mechanisms to find groupmates.
4. Continue using your document from lab 7 if you have already started it. *(BEGIN same as lab 7)* Otherwise, you should then create a shared document that everyone can work on together. Please clearly indicate group members on your single submission so we can appropriately credit everyone. For a question marked "Team", you should have one block of text from the whole team, while for a question marked "Each", each team member should identify themselves and provide their own block, following a format like the following:

Task 2A

<team answer here>

Task 2B

...

Task 4B

Task 4C Alice

<Alice's answer here>

Task 4C Bob

<Bob's answer here>

Task 4C Carol

<Carol's answer here>

(END same as lab 7)

Your Tasks

- **(TURN THIS IN, 5 points)** First, read the assignment specification and ***estimate how long you think it will take you*** and write it down.

1. Pre-task: Data and explainable classifier preparation. Follow the pre-task to understand how we set up the data and explainable classifier.
2. *Task 2: Explanation Consumption:* These instructions are duplicated in the notebook, use the code in the area to answer the following questions.
    A. **[Team, TURN THIS IN, 5 points]** Sensitivity-based explanation varies features one-at-a-time. Suppose we varied features in a pairwise fashion. How might the output vary?
    B. **[Team, TURN THIS IN, 10 points]** Compare and contrast the explanation output on the various sample groups (see cell title "Sample Groups").
    C. **[Each, TURN THIS IN, 10 points]** In what ways and to what extent do you think these decision processes (as implemented by the raw and processed model) are biased? Compare and contrast the classifiers trained on raw and processed data. Consider the evidence in both the explanations and validation notebook for this question.
    D. **[Each, TURN THIS IN, 10 points]** Which pieces of evidence do you find most compelling in reaching the judgment you did in the previous question? And why are those pieces of evidence the most compelling to you?
    E. **[Each, TURN THIS IN, 5 points]** There is a tradeoff between having brief global descriptions and having meaningful local explanations (global explanations have a length scaling on number of BOTH

features AND categories within each feature, but a low dimensional feature/category space means fewer perturbations are possible for sensitivity-based explanation and also that the distances used for case-based explanation are more likely to be small). Reflect on this tradeoff, particularly as it pertains to the amount of text you would like provided in each explanation.

F. **[Each, TURN THIS IN, 5 points]** Which explanation strategy do you like most and why? And which explanation strategy do you like least and why? (If you want my opinion on Task 2F, consult the following workshop paperLinks to an external site. I wrote on the topic. Your opinion being at odds with mine is fine; provided you have a strong argument for whatever stance you take.)

3. *Task 3: Big Picture*
   A. **[Each, TURN THIS IN, 15 points]** Consult the following short paper by Baeza-YatesLinks to an external site. (especially figure 1) and criminal justice process figureLinks to an external site.. and provide any reflections comparing and contrasting the works, particularly the figures.
   B. **[Each, TURN THIS IN, 15 points]** Wade into the holy war by skimming a bit of the following sources, in which ProPublica (journalists) and Northpointe (the software vendor putting these decision support systems into courtrooms) argue about whether or not the system is problematic. Which arguments do you find compelling and why?
      I. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithmLinks to an external site.
      II. https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysisLinks to an external site.
      III. http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdfLinks to an external site.
      IV. https://www.propublica.org/article/technical-response-to-northpointeLinks to an external site.
   C. **[Each, TURN THIS IN, 15 points]** Having finished this assignment, re-read what you wrote for task 1A in Lab 7. How has your position on the use of this kind of technology for this kind of problem changed?
4. **[Each, TURN THIS IN, 5 points]** Upon completing the lab, determine how long you actually spent on the lab, and report that timeframe in addition to your estimate beforehand.

Submit

One submission per team should be fine, but be sure to indicate ALL team members on that submission.