

Co-creating learning experience with students in group data science projects

Qian Zhao

Department of Biomedical Data Science, Stanford University

Project Setting

Your project group is working with a school district to design a part of a new choice-based student assignment policy, i.e., students rank schools according to their preferences, and if the number of applicants exceeds available seats, a tiebreaker will be assigned to break ties. The district aims to assign a higher priority to students from historically underserved populations to provide equitable access to education. Because of privacy concern, the priority can only be assigned to a residential block rather than an individual student (i.e., every student living in a block will either be given the priority or not). In this project, you need to determine how to pick which residential block to assign the priority to.

I “What do we need to learn?”

When you facilitate discussions, guide students to think **critically** and **deeply** about the project goal and their intended deliverable. Invite students to brainstorm approaches, and as they do so, ask them to think about what additional concepts or knowledge they need to learn. This is the starting point of **self-directed learning**.

You can use these questions:

- What do we want to achieve at the end of the project?
- What else do we need to know about the project?
- Any thoughts or ideas (on how to achieve ...)?
- What do we need to know to be able to ...?

Step 1: Guide students to think critically about the problem at hand and assess objectively their own knowledge and skills.

II Handout/Tutorial

Suppose each subject has a true label $Y = 0$ or $Y = 1$ and a binary classifier assigns a classification of either $\hat{Y} = 0$ or $\hat{Y} = 1$. The table below shows four possible combinations:

	$Y = 1$	$Y = 0$
$\hat{Y} = 1$	True positive	False positive
$\hat{Y} = 0$	False negative	True negative

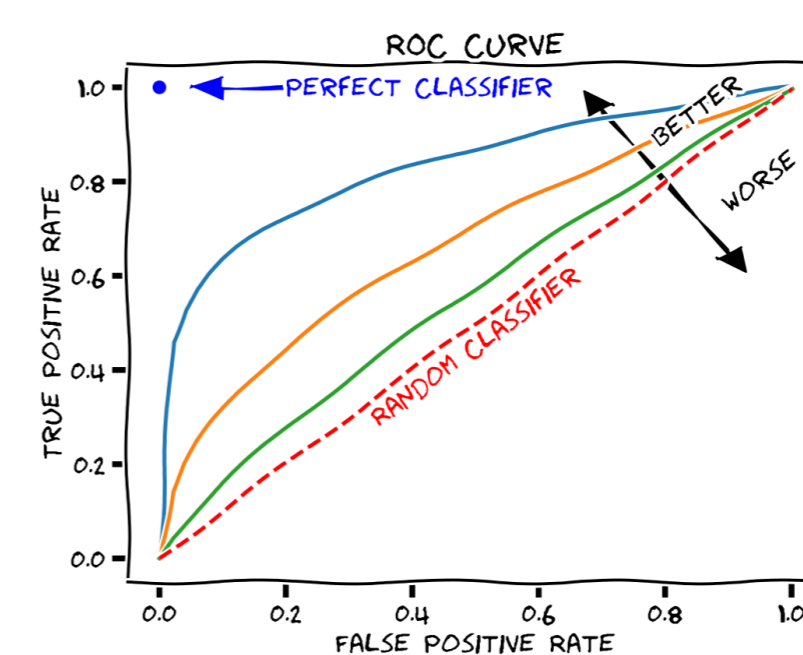
There are many criteria one can use to assess accuracy of a classifier:

- Accuracy $P(Y = \hat{Y})$.
- True positive rate (TPR, recall, sensitivity) $P(\hat{Y} = 1 | Y = 1)$.
- False positive rate (FPR) $P(\hat{Y} = 1 | Y = 0)$.
- True negative rate (TNR, specificity) $P(\hat{Y} = 0 | Y = 0)$.
- False negative rate (FNR) $P(\hat{Y} = 0 | Y = 1)$.

Step 2: Provides resources (e.g., handout, tutorial, lecture notes, videos) for students to learn individually. You can also work with students to find resources.

III Exercise & Discussion

1. What are the true label Y and the classification \hat{Y} in our project?
2. Can you draw the ROC curve for one approach to assign priority?
3. Any thoughts after reading the handouts?



Step 3: Provide deliberate exercises, and guide students to connect what have learned with their project work.

IV Generate ideas & approaches

Now that students have practiced these concepts and skills, it's time to apply their newly acquired knowledge to generate new ideas!

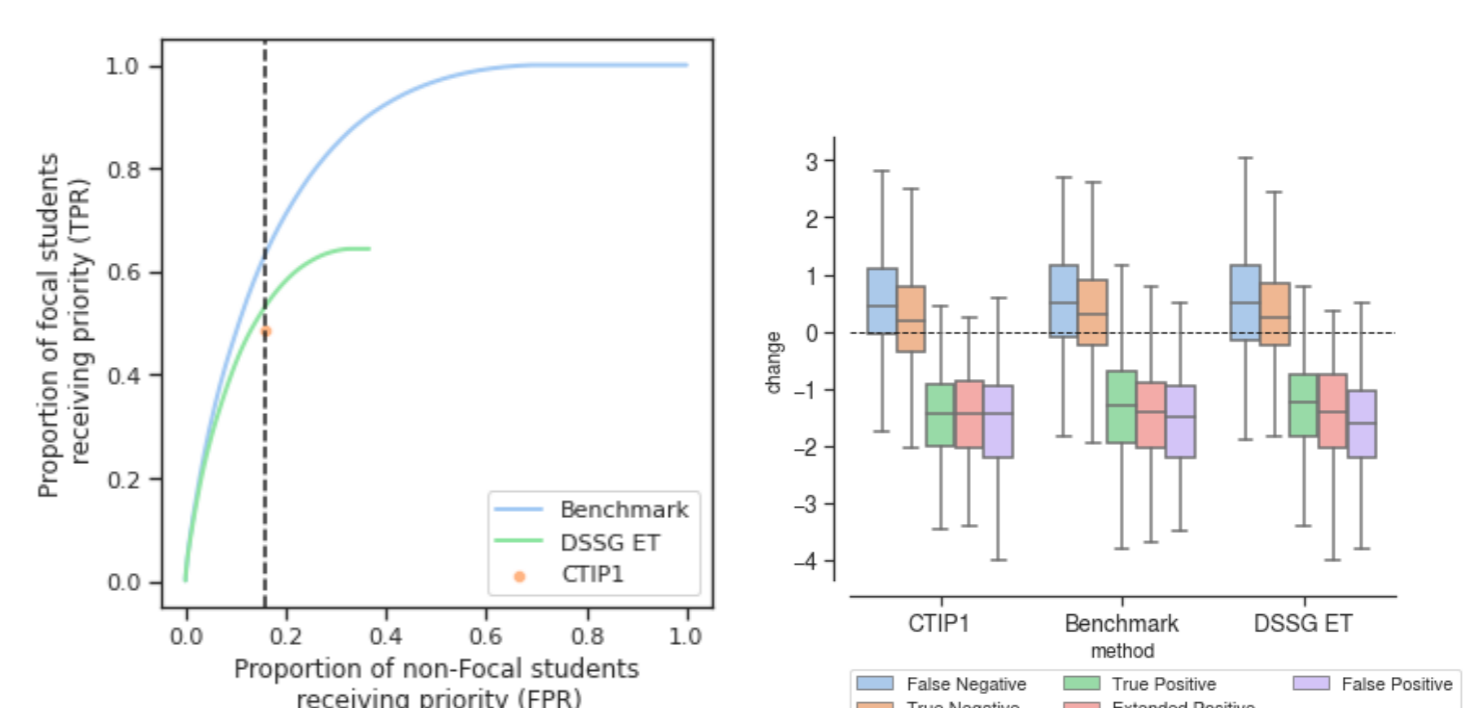
You can ask these questions:

- Any thoughts on how we can ...?
- What can we try first?

Example We can assign priority as follows: for any number of students to give priority to, we choose residential blocks by optimizing the number of “true positives”, i.e., the number of students from historically underserved populations.

Step 4: Work with students to brainstorm new ideas! The new concept, knowledge, or skill might be used to frame the problem, motivate a method, or inspire an approach.

V Project work



Step 5: As students work on the project, facilitate the team to discuss progress and come to a consensus.