P52 3:08

Let's see it

P50 3:11

cool. Okay, so before before I move on again, any other questions? No? Cool. Okay, have a great time. Thanks.

P52 3:20

Thanks.

P51 3:22

Okay, there's some credit Google Doc.

P50 3:25

Yeah, I will create a Google Docs over.

P49 3:29

I could, um, I can throw my screen up so we can look at the models together if you guys want.

P51 3:36

Yeah, that works. Okay.

P49 3:49

So I wanna, I think we should start with the overview so we can see just kind of what's good and what's bad in each. So it looks like model one is only good at accuracy as as much Model three or it's like all relative, right? So I, I consider if it doesn't have an x then it's probably pretty decent. But we'll see if that holds true. Model two is good at false negative rate and then the disparity in accuracy, but bad at actual accuracy and bad at FPR. Model four is only good or really good at disparity in accuracy. Five is good at false positive rate and disparate error the yeah the disparity and false positive rate but bad at false negative rate. Six is interesting because it's bad at disparity in false positive and false negative rates. Seven is good accuracy and false positive rate but bad at false negative rate. An eight is Bad accuracy false positive rate and disparity and false positive rate, but it's good at false negative rate and disparity of accuracy. We'll do this we'll start from the top. So who here

**P50 5:18**

can talk about our roles? Like which? Oh, yeah. And then maybe like, what model? Do we choose as the choice? Yeah, I can start. So I think the persona that I was given it's a community member. And according to the car, my goal is to reduce the false negative rate because I don't want those real offenders or criminals hanging out in my community. I think my my first choice is definitely model eight seems it has the lowest false negative rate. So if I remember correctly, the false positive rate from our aid is around like 15% But as we can see also like, yeah, yeah, if you go to that car that would be nice. Yeah, yeah, always like 80% all the way has like a, like a 10% disparity between why American African American, but overall is pretty low. And then I think my second choice model two, which has the second lowest part false negative rate, but I think like the pros for model two is that he actually has a pretty good disparity especially because it's disparity accuracy is zero. And also this disparity in false positive and false negative rate is also lower than the ones from the model eight. Um, so I will say if my only goal if I only have one goal and that goal is to reduce the false negative rate, then I would choose model eight. Before more balanced model, I would probably choose model two.

P51 6:54

Got it?

**P49 6:58**

Well, I've been in I was I was a judge. So my goal was to have a higher accuracy rate overall. So the ones I really looked at were one, three and seven. Because they were the kind of the cream of the crop when it came to accuracy. So it's interesting though. So I actually ended up picking as my first choice, the one with the highest overall accuracy. So model three here. And that was kind of bolstered by everything else. So we see a lower disparity in accuracy here, then model one, model seven, I kind of ruled out because of the how high the overall negative false negative rate was. So I said, hey, that's not good as compared to models one. Well, okay, one's not great either. And threes, threes, D But so yeah, I was looking at model three and model one mainly. And I picked model three as my first choice just because overall, I think I saw a better false positive rate here or okay. No, I didn't.

Why did I okay, so it was the overall accuracy for sure. And then I think it was disparity values. So I saw lower disparity across the board.

I think that's why I chose model three over model one.

**P51 8:37**

Yeah, okay. I can go on. So I was a defendant. So I'm really worried about being falsely predicted as like a repeat offender because I don't want to go to jail, you know, just for no apparent reason at all. So I, I really want a a model of a lower false positive rate. So I my first choice was model five. Yeah, so it is a super low false positive rate. And the disparity of all positive rates across the two groups is also really low too. So I really like this one. And I mean, accuracy isn't bad either. It's pretty good overall, a bit of disparity, but obviously, there's a low cost is very high false negative rate. To me, I didn't really care about that, because I actually went don't want to go back to jail. Um, and then the second model that I picked was vol seven. And this one is a bit better well rounded even though that the second lowest false positive, it's pretty good on accuracy, disparities a bit and, but it also has a lot of false negatives. So I mean, this one to me, I can live with a slightly higher false positive rate while the model is actually better overall, and I think I'm like, in terms of how this affects me, I think, a false positive rate of 10% or Lower I'd probably be happy with but obviously this is the nice, low 5% for the false positive rates actually a lot.

P49 10:10

Great. And your first choice was you said five and then it was

P51 10:14

five and second was seven, but I'm now kind of leaning towards more seven just because it's nicer overall. Yeah.

**P52 10:23**

Okay, so I'm a fairness advocate. So what I'm looking at is like minimizing the difference between different demographic groups, Milo catala, at the disparities between the different demographic groups, and so I first took a look at four to four or five and eight because they have like stars in the disparities and stuff. And my top choice is to actually because it has like pretty good disparity for accuracy which is zero. decently low disparity for false positive and false negative rates, although its accuracy is not very high. But yeah, I want to focus more on the fairness between the different groups. And my second choice is model five.

Yeah, model five.

Also has like a pretty low disparity in like accuracy, false positive rates and false negative rates, or, yeah, it compared to like foreign aid. I think five is better in the false negative rates and yeah, false positive rates. So, model two and five are like more fair for the different groups, although like model five has, overall a very high false negative rate. But it's I'm pretty close between the different demographic groups. So I'm actually not very sure about that consequence of like having a very high false negative rate for both demographic groups, but it's pretty similar. Similar between the group's Yeah.

P49 12:24

Yeah, so I heard seven come up twice. Five came up twice. Did we have any other overlap?

Yeah, two, two came up as well. Okay.

I didn't actually pick seven but I took a hard look at it. But honestly, it has the second highest overall accuracy. So maybe I should have.

P50 12:56

Um, so actually, I think like accuracy is probably not Good metrics in every cases, I would agree there is still like a bias in the training data set. Like for example, if 90% of the data is like is defined as positive case, or 99% to the extreme cases, then you move the threshold to the very right, you identify everything as positive, then you will still have the 99% of accuracy. Um, so, I think, yeah, I think we can discuss, like between the three models and maybe eventually one

P49 13:31

so, so what are the three that we're taking a look at model seven, model two, and what was it five,

P50 13:38

Yamaha two, five and seven. Cool.

P49 13:45

So what I'm gonna do is

actually, I know you've got one going over there and but I'll make one word so we can take a look at all.

All of them at once.

They're gonna be out of order. I apologize.

P51 14:18

I mean, I can just I can start off saying like, if it's either five or seven, I'm good with that, you know, because they both have low positive rates. And that was both my choices.

P49 14:29

Yeah, I was about to say you got both of yours in there. Yeah. So,

P51 14:32

hey, do we pay five or seven?

pretty happy with the way.

**P50 14:38**

But I think from my perspective, like I was 17. So like a really bad model because it's pretty high. It's like 87, like around 80 or 70. So I will probably choose model two. If you took them off, take a look at Moloch he has like 40% of poverty and hunger, South Boston. It's like a group balance and accuracy. I wasn't I was not very good, but I think it's I think it's also not very bad. So it's something like romney right in the middle. Um, yeah. If any of the Compare like mala to tomato seven, it doesn't it has I think Monica has an accuracy of 60% while seven it's like oh 9% above that. And also like in the disparity in terms of dyspraxia perspective, you also have like a zero inaccuracy disparity. And then the false positive and false negative is disparate, always for under 10%. So that's why I am advocating for model two.

P49 15:42

So I think, as a judge, I'm pretty happy if we have accuracy at 60% or over and we hit that across all three of these. I don't know what a judge's second metric would be. But I don't know in the American legal system I think false negative rate is considered like you know innocent till proven guilty so hopefully that's a pretty big metric but or false positive Excuse me. So that makes me feel like either model seven or model five but not model to

Dylan and I are pretty close in our in our agreement then I think

**P52 16:40**

Yeah, he has a fairness advocate i would i would go for it like two or five over seven because seven has like much higher disparity between the different groups. Yeah, although like to has like, slightly lower accuracy and stuff, but two is definitely the best for, like fairness between the groups. And then one, like too much difference between white Americans and African Americans in using the, the algorithm. Yeah.

Huh. tricky.

Especially from our second model. Yeah, go ahead. Sorry. Yeah, I was just gonna say like for mortal seven, the disparities in the first negative rates is like 15, which is much higher than both two and five. In that, yeah.

P50 17:53

Yeah, I think from my perspective, if you guys like if we choose modify, I would, I would Listen, I would panic because the false activates that. Yeah.

P52 18:03

The um

P49 18:12

Oh, lost you don't

P52 18:19

know No.

P49 18:22

Oh, there you go. Sorry, we lost you there doing

P51 18:26

all right, sorry.

I'm cleaning through a model three. I think I'm looking at right now actually. And I think it might not be a bad one to meet in the middle a little bit.

P49 18:37

I like the way you're talking.

P51 18:39

Yeah. Because I mean, it's I'm taking a higher false positive, but it's a definitely a big step lower and false negative. And I mean, the disparities aren't that high. And also, it's pretty accurate to maybe I should have done like,

P49 18:51

slides and stuff.

P51 18:57

I mean, I think obviously, this is a bit of a better choice than doing model like five where the false negative rate is like 90%.

**P49 19:07**

So let's discuss that. Do we do we all deem that high of a false negative rate as unacceptable? Is there anybody who's okay with the false negative rate being that high?

P51 19:22

Um, I think I would be okay with I mean, obviously, I'm, I have a very, very nice false positive rate, but I think there's anything societal here. Like, it's definitely not the greatest thing to have a false,

P49 19:34

right. And that goes into like, the weight of each of these metrics, right? Like how how important they are.

P51 19:42

That's gonna have a large impact on the community if that false negative rate isn't 90%

P49 19:46

I agree.

Is everyone okay with us eliminating model five?

P52 19:55

okay with it. Okay.

P49 19:59

All right. Let's uh, Let's eat it.

All right.

I guess I should ask, are there any metrics in any of these models that people feel are unacceptable? Like we had that with that false negative rate?

P51 20:19

I'm just a little bit worried about the false positive rate in mouth not to say I agree with that the percent that's really bad for us defendants

P49 20:29

kind of insane.

**P51 20:31**

Yeah. And you've already like, entered the, the, the word terror, being being locked up in jail once and having a 50% shot at being predicted of being under badly, like in a bad sense, like where it's not actually going to happen. That's just I can do so

P49 20:56

saying if you felt strongly about about Model two as like a good thing. Can you talk us through how you justify that false positive rate?

**P50 21:07**

Um, I think I'm bringing model three. Um, I mean, like, I think like, we should all consider other people's perspectives, as well as a defendant will also consider or what is good for the entire society. Then after listening to what he has talked about, I think it's it is bad for us to have like a 50% of false positive rate, because those people will have already gone to the Joanne's and then they have like 50% of the chance to be put into the jail again, even if they are, you know, bent. Well, so I think maybe at that point, um, I don't think that's the optimal choice like model to, I'm willing to sacrifice some of the false negative rate for like, lower false positive rate for

P49 21:59

a reason Are we good eliminating model to Mr. fairness advocate? Is that okay, we're getting rid of your perfect disparity.

P52 22:07

Yeah, I think looking at the disparity for model three, it's, it's okay, like, yeah, false positive and false negative, it's pretty similar to, to to also, it's just a little bit higher in an accuracy. But I'm still a little worried about like the false negative and in model three, as compared to like, two. I don't know how to, like it's actually lower than 2.2%. The false negative rate,

P51 22:40

oh, the rate in general. Yeah, but to make this better.

P52 22:44

No, I was. I was thinking about that, like the higher false negative rates.

P50 22:52

I think model fora

P52 22:55

okay. Yeah.

P49 23:01

So model four has very good disparity in accuracy and pretty comparable disparities elsewhere. almost a third false positive still.

P50 23:16

Yeah. Um, I think in this case like false positives and false negatives kind of like

P49 23:23

each other. So I guess it comes down to the question which is more important. In you all's mind is a is a false conviction more important than a false freeing, or is a false freeing more important?

P51 23:41

This is one of the hard the ethical debate. Yeah. Yeah.

P49 23:51

In my mind, just like outside of the being a judge, I feel like a false negative rate or a Excuse me, the false positive rate is actually much more egregious, like,

that feels very autocratic and very

unhealthy for society as a whole to have almost a third, wrongful imprisonment. Especially when you do have a discrepancy between races here as well. I think that might not be the best in terms of

**P51 24:3**4

Yeah, I'm in agreement on that. Obviously, just because it's a pretty high false positive rate still in granted, it's lower than 45% from model think model two was but right. But it's still not the greatest in terms of what I would want. Yeah. The third being wrongly wrongly convicted again, right? That That's not good. Yeah.

P49 25:09

So if we had to get rid of either model for model two, which one would you guys get rid of?

P51 25:23

Ma? I mean, I would say goodbye.

personally.

**P49 25:29**

I am in agreement. Yeah. Okay, so get rid of model two. And we want to add, should we add model for? Or? I mean, how's everyone feeling here? I think Dylan and I agree that we don't like that false positives still, when compared to the other two false positives. But we also see higher disparity here, I believe, is that correct? Yeah. So disparity and accuracy. is much lower here than these other two but

reasonably comparable

FPR and fnr to model three

and the false negative

much lower.

I don't know. I don't want to steamroll I

I don't really care for model four personally as a judge

both of these have higher accuracy as well.

Across the board I believe

that true Yeah, yes, yeah. Across the board higher accuracy on both model seven and model three.

P50 27:02

Yeah, I think in my case, I don't care about accuracy that much. Right. Um, I think I probably like she's building model for your model for

P49 27:11

Oh God, where are we supposed to have our proposal written?

Oh, no. Okay, quick snap decision. I like model three personally.

P51 27:24

Yeah, I'm doing.

P52 27:26

I would also choose between like four and three. I eliminate seven.

Unknown Speaker 27:30

Yeah. Okay.

P49 27:35

I say we're good with model three.

P50 27:39

Yeah, it might go with model three, but okay. Yeah, that's data. Um, and what about the proposal?

P52 27:46

I don't know.

P50 27:50

Um, I can quickly write about it.

P49 27:53

Is that okay? I feel really bad.

P50 27:55

I think I think that's fine. That's fine.

P49 27:57

Okay. Thank you.

I buy you a beer when we can meet in person Holy shit. Yeah.

P50 28:06

Yeah. Cool

P49 28:07

in February. All right by.

Unknown Speaker 28:16

Right Welcome back.

Unknown Speaker 28:19

I hope you guys have fun discussing and negotiating with each other.

P52 28:23

Um,

Unknown Speaker 28:25

so now we're gonna wrap up.

Unknown Speaker 28:30

So a couple to do's for you guys.

Unknown Speaker 28:33

The most important thing is to follow the instruction to the end in the in the assignment on campus and actually click on the Google Drive link. So when you click online, you should see something like what I'm seeing right now, basically a Google Drive. So one of the documents called team proposals, so you should click on a pin proposals. The person who is responsible for writing a document or maintain document, you'd put the link here. So I see already two of the rooms already put the link here, which is great. Another thing is the person who is responsible for the recording should actually upload the recording to your

Transcribed by https://otter.ai