Unknown Speaker 0:02

Awesome. So yeah, just follow the steps,

Unknown Speaker 0:03

everybody. Okay with the instruction on the canvas, everybody?

Unknown Speaker 0:09

Yes,

Unknown Speaker 0:10

yeah. Okay. So just follow the steps is very clear. Yeah.

P24 0:16

Thank you.

Unknown Speaker 0:17

All right, thanks. Thanks. Bye.

P21 0:21

Okay, so I guess let's just start with everyone's top two models. What did you pick?

P22 0:29

I picked model fo and the model three models two and 34343.

P21 0:39

Okay, let's see. Actually, I don't know if we Um, I think we all have the same models but I don't know. So okay.

Why did you go with four three?

P22 0:56

Well, then he let me recheck.

P21 0:59

Yeah, it was

P22 1:00

I can see my faults with overall accuracy about 64.1. And the reason I choose it is I think the false positive is more serious than false negative because innocent men into jail so, one of my rules is the false positive rate and the negative rate should be cannot balance and it will be okay if the post the positive one is slightly lower than the negative one.

P21 1:38

Accuracy

Unknown Speaker 1:40

low

P22 1:48

and also

the four three

P21 1:58

Okay, so

mileage was a little different. I think I went with

let's see, I think it was two and yeah. Yes, it was.

It was to an eight. I think. So.

I don't know if that makes me draconian or whatever. But, so I was actually prioritizing low false negatives. So, and my reason is, this isn't like this isn't like a judge AI isn't deciding guilt immediately. It's deciding, Hey, is this person likely to commit crime again? Do we need to watch them? Do we need to just

you know, keep more of a close eye on.

P22 2:56

I see to be more strict.

P21 2:59

Yeah. So I think it actually should be a little bit.

Unknown Speaker 3:04

I mean, assuming

P21 3:08

Okay, budget, I guess is the fact that you can monitor everyone. Well, you can. That's like China, but most countries do not have the budget to monitor everyone. Um, well the resources. But still I think that

Unknown Speaker 3:28

like,

P21 3:31

basically I'm trying to say is I'm trying to prioritize public safety on one hand.

I also wanted to prioritize

low what was it What was the discrepancy? disparity disparity? I want to prioritize low disparity between demographics so and basically the reason for that is I think that AI should be, you know, judiciary AI especially should be impartial about that sort of thing. Shouldn't like, you know, be horribly horribly biased towards one group or another in terms of like being more accusatory of one group over the other. So, across the board, I think low disparity between demographics is really important for this sort of system.

P23 4:22

I can go next slide. I missed the first part, but I yeah.

P21 4:29

Which wasn't you pick.

P23 4:31

I picked five and seven. So for me the reason that is model five. For me, the reason was I was basically looking at false positives and true negatives. I mean, whatever it was, I'm forgetting the terminology. But basically, with model five and seven, the false positives were low, I think one was 1% and the other was 4.7%. So in that sense, I thought that This, in this in these two models, it's less likely that a person will be falsely accused or be not accused, but falsely thought of doing a crime is less likely, I mean, it's less likely that they will be falsely accused of maybe having the inclination to do another crime. And at the same time there one was false positive other was true negative. Their true negatives were also very high. I think one was 90%. And the other was, I don't remember the numbers. And those were high. So I thought that the algorithm is actually predicting it correctly, that that all the these people were rightly predicted that they were even if they did one crime or the algorithm today it correctly that they are more likely to do the next one on not likely to do the next one. So the accuracy rates were higher there. And those were my reasons for selecting model five and seven. And at the same time, the discrepancy between the two groups for both model five and seven was much less compared to probably all the other models, because model five and seven both has had like a very small difference in false positives and false negatives. So in that sense, I felt that the algorithm is not biased towards one or one or the other group as well at the same time.

P21 6:43

Right. Okay.

I didn't know we're supposed to do with the Google Doc afterwards, but

P23 6:57

do we have to agree on one model or Do we just have to do a

P21 7:01

good point we have to reach consensus, because so far we've chosen all different things.

P23 7:07

Again, somebody Canvas wanted.

P24 7:11

We don't have to reach consensus. We just have to explain.

P21 7:14

Okay, wait, now, did you get the Google Doc link or no?

P23 7:19

Oh, yeah, I already got the Google Doc in.

Unknown Speaker 7:21

Okay.

P24 7:23

Awesome. Yeah.

P23 7:25

I never, I actually didn't go to the part where they talked about persona or whatever. I never, I don't know. I don't think I got that. I don't think I just the bottom line directly to models. And I looked at models and did the Google form that said, nothing between quiz and the Google Form.

P21 7:43

Yeah, I don't know if this. That's the thing though. We don't have personas, but some people do. So I feel like I'm hoping that doesn't like void the experiment that we know some people have personas. We didn't but I don't even know

P23 8:01

I can go and check with in the main room. It's

P21 8:06

probably fine. It's probably fine. I mean, if the President mentioned it, it's probably not insane. Um,

**P24 8:14**

so I can just discuss my two I went with actually went with one in three. The reason that Brittany was was, I think, just for a lack of context, because we're This is, I think basically prioritizing, just conviction versus non conviction, right, but it doesn't actually go into any context regarding what the criminal act was or wasn't. So without that extra context, I think it's hard to determine whether false positive or false negative is more valuable. And so I went towards the ones that are generally just yield the highest accuracy. And so that was my main rationale. and choosing the

P21 9:13

I mean, the public safety argument doesn't really work. It's like,

P24 9:17

I don't know,

P21 9:18

if the crime in question is jaywalking. Right, then I, you're probably right at the public safety argument doesn't really work in that case, and we don't really have context. So that makes sense to me.

Unknown Speaker 9:33

Okay.

P21 9:35

We are supposed to read a short proposal.

Oh, it says okay, if we're if we reach a consensus, we make a proposal. Otherwise, just share our reasoning. Okay. Okay. So I don't think we I don't think we can make Kindi thing that you could turn as content up I mean, other than maybe like prioritizing accuracy, which I think it's hard to not want to prioritize. Oh, hi, professor. I'm recording to this room.

P23 10:16

I think Yeah, curacy can definitely I could if he can definitely be one of the factors. But yeah, I'm not too sure if we can still think of reaching the consensus.

P24 10:30

Opinions change since we've shared?

P21 10:32

Um, I mean, mine have a bit. So my false.

My argument about the false negative rate was mostly predicated on the idea that these systems aren't being used for conviction so much as just for recommendation. Um, you can't just have an AIB The budget court case, right? Um,

P24 11:10

well, because I mean, on that one, the question that comes in 60% enough accuracy for them actually use it reasonably?

P21 11:17

I mean, that's true, but the accuracy, like is between 60 and 70%. Just across the board. So

P23 11:27

41% actually true. Wasn't that accurate? Wasn't the accuracy high? What's the accuracy for five and seven? Five

P21 11:39

is 66.7% accuracy. And seven is 69.1%. accuracy.

P23 11:46

Yeah, I think in terms of accuracy, it was approximately that percentage for everything.

Unknown Speaker 11:52

Yeah.

P21 11:55

Also looking at model number two, I don't know if I favorite anymore because it has the lowest accuracy of any model. But the accuracy is the same for white Americans and African Americans, thus the disparity is zero it's equally bad. So I don't know if that's a good thing.

P24 12:12

Um

P21 12:16

Yeah, no, these models mostly differ on the false positive and false negative rate. Like some like um, but number five has a really high false negative rate. They're really low false positive rate. Whereas

Yeah, no, these mostly trend towards having higher false negative rates than the higher false positive rates anyway.

Other than maybe like number two?

P24 12:58

Yeah, I think if I had to pick a bad I'd probably go with model three. But

P21 13:04

I mean yeah, it has the highest overall accuracy that seems simple enough.

And it also has a decently low disparity 2.1%. Actually,

it has other than monk who has the second lowest disparity of all the models, or four has the lowest. Yeah, it also has a low false positive rate. So if we want to err on the side of leniency three is probably the best option overall.

I mean, I don't mind that but it's really just a matter of

we have so so little context about how the system is being, how the system is, and what the crimes are that being trained off of and just

P24 14:00

When everyone's thought,

P23 14:00

what are the numbers for? For model three, four false positive

rates overall,

P21 14:08

the overall false positive for three is 10.1%.

P23 14:12

And how about the one in the bottom right quadrant,

P21 14:15

for those negative is 65.2%. With a pretty big disparity between white Americans and African Americans, there actually, is a good amount of as a fairly large disparity overall, I think like, I mean, what the false positive rate is triple for African Americans and it's low overall.

P23 14:39

Yeah, I think I think those are my reasons to go for five because five has the lowest had the lowest disparity. And in terms of overall false positives and false negatives, the numbers were at least I think, 1% and 90%. So at that point, there was more I mean, less disparity between the two groups and more. I mean better rates.

P21 15:07

I think we can all at least agree that model six is just

I think that's the only one.

Yeah, that's the only one choice is model six, because the disparity

P24 15:19

is absurd.

P21 15:20

Like it hasn't doubled the full 40% false positive for African Americans.

P24 15:29

I temporary can't like picking an agreement to just write down rationales for models.

P21 15:35

Yeah.

P23 15:36

I mean, they're both supposed to write those. Oh,

P21 15:40

that's what I that's why I asked about the Google Doc.

You have that right.

Unknown Speaker 15:48

No,

P22 15:49

yes. I think we have the cool dog.

P21 15:52

Okay. I don't know she's out of the room. One.

P23 16:05

I think I got disconnected. Do we have to write these individually or is that one group submission?

P21 16:10

I mean, it's one group submission and I have to duck for it and yeah,

P24 16:18

yes, we have to wait. Just like a paragraph that

we discussed. All right.

All right.

Unknown Speaker 16:57

So um

P24 17:05

How's it is

we could just say that we had no basis on which to prioritize falls.

We can just say like reach beyond our reach. I like her own personal reasons. Some people like value to false positive, higher false positive versus others are sorry lower false positive versus others.

P21 19:30

Is my keyboard audible?

Did you hear that typing

P24 19:36

is not listening.

P21 19:37

Okay, well we'll see if it shows in the recording. It just occurred to me I'm using a mechanical keyboard and that might be loud

Unknown Speaker 22:13

Looks good to me.

P24 22:16

Yeah, that works. Okay.

P21 22:21

Let me check what's supposed to be next on the list. So,

P24 22:27

proposal done.

P21 22:37

I think we're good to go.

P24 22:43

Okay.

P21 22:44

So I'm gonna end the recording.

It hasn't been lovely. We're going to

Transcribed by https://otter.ai