

Evaluation of WIT using the Guidelines for human-AI interaction design

From Gorgin and Xin

Severity*

The following 0 to 4 rating scale can be used to rate the severity of usability problems:

- 0** = I don't agree that this is a usability problem at all
- 1** = Cosmetic problem only: need not be fixed unless extra time is available on project
- 2** = Minor usability problem: fixing this should be given low priority
- 3** = Major usability problem: important to fix, so should be given high priority
- 4** = Usability catastrophe: imperative to fix this before product can be released

Source: <https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/>

Issue 1: [changes unclear]

Heuristic:

[G4. Show contextually relevant information.]


Description:


[Unable to see what changes are caused in „Performance“ by changing/setting parameters.]


Severity*: [3]

Recommendation:

[Option 1: Highlight the block where changes are caused;
Option 2: Generate an window to show user where changes can be found]

race  Shows the model's performance on datapoints grouped by each value of the selected feature.

Slice by (secondary) 

<none> 





Fairness

Apply an optimization strategy

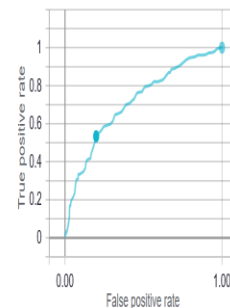
Select a strategy to automatically set classification thresholds, based on the set cost ratio and data slices. Manually altering thresholds or changing cost ratio will revert the strategy to 'custom thresholds'.

- ☐ Custom thresholds 
- ☐ Single threshold 
- ☐ Demographic parity 
- ☐ Equal opportunity 
- ☒ Equal accuracy 
- ☐ Group thresholds 

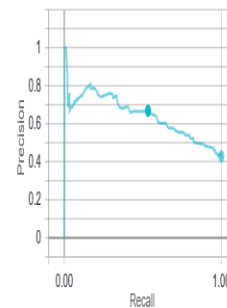
Equal accuracy thresholds for 6 values of race

Feature Value	Count	Threshold 	False Positives (%)	False Negatives (%)	Accuracy (%)	F
African-American	1904	 0.52 	23.2	12.2	64.7	0.70
Caucasian	1111	 0.47 	11.3	20.0	68.7	0.59

ROC curve (AUC: 0.71) 



PR curve (AUC: 0.64) 



Confusion Matrix 

Issue 2: [Index unclear]

Heuristic:

[G4. Show contextually relevant information.]

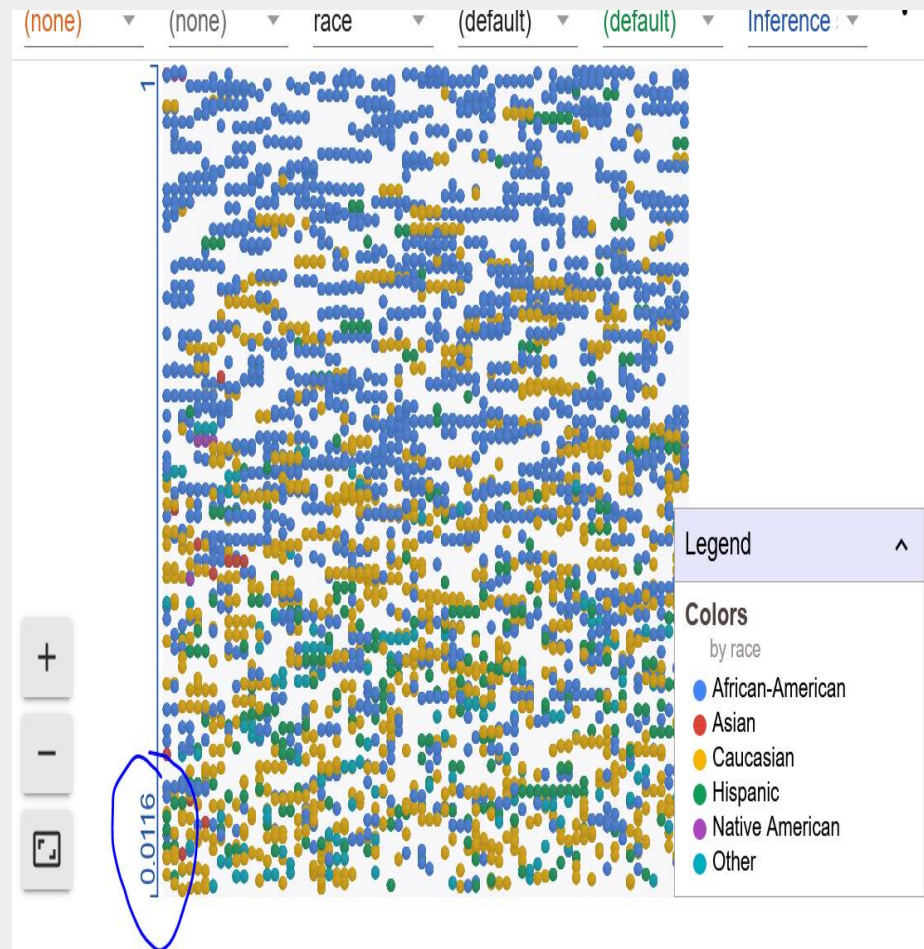
Description:

[At first sight it may confusing why there is a 0.0116.]

Severity*: [3]

Recommendation:

[Just add information of 0.0116 which informs the users that the possibility of all data points of recidivism ranges from 0.0116 to 1.]



Issue 3: [Variable/Attribute unclear]

Heuristic:

[G4. Show contextually relevant information.]

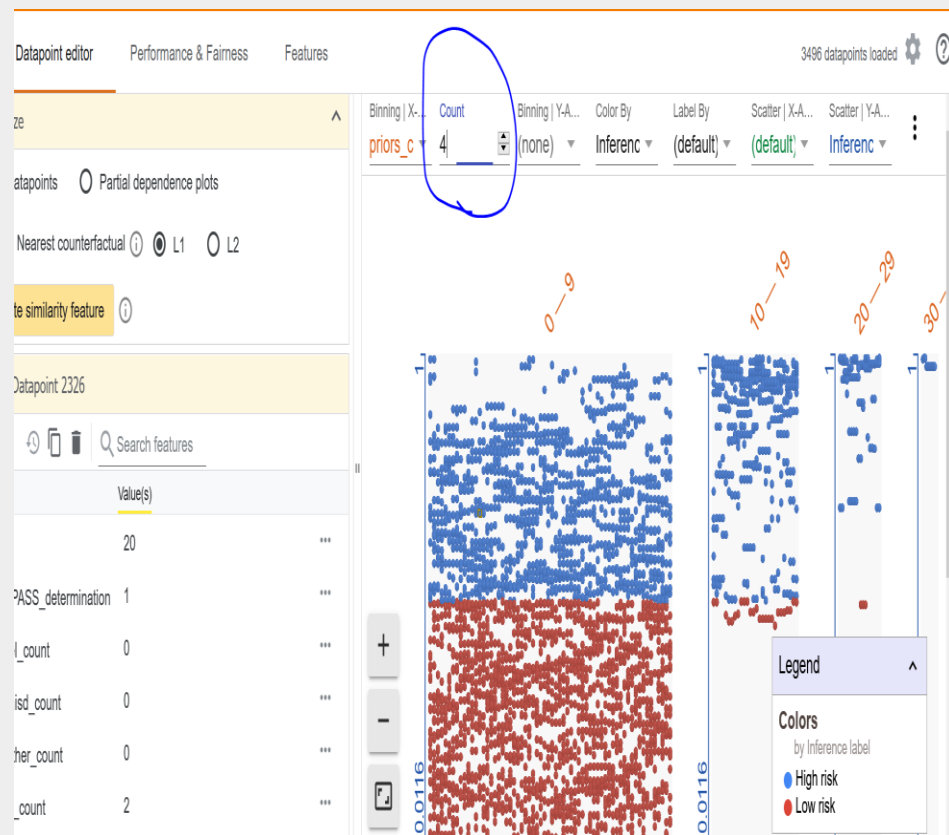
Description:

[For some attributes, it is unclear what the „count“ means and influence the display.]

Severity*: [3]

Recommendation:

[Add additional information.]



Issue 4: [L1 & L2 unclear]

Heuristic:

[G4. Show contextually relevant information.]

Description:

[L1 & L2 are not clarified and users may don't know what are they and how to choose between L1 and L2.]

Severity*: [3]

Recommendation:

[Tell the users how L1 and L2 regularization will influence the result of nearest counterfactuals.]

Visualize

☒ Datapoints ☐ Partial dependence plots

☐ Nearest counterfactual i ☒ L1 ☐ L2

Create similarity feature i

Edit - Datapoint 2085

< > ↺ 📄 🗑

🔍 Search features

Feature	Value(s)	
age	25	...
COMPASS determination	0	...

Issue 5: [no “undo” option]

Heuristic:
[G12. Remember recent interactions. G9. Support efficient correction.]
Description:
[There is no „undo“ option or checkpoint preservation to go back to a previous status.]
Severity*: [3]
Recommendation:
[Maintain short term memory or give users the right to preserve their own pathways.]

(optional) Screenshot