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

From ansa and mane

Severity*

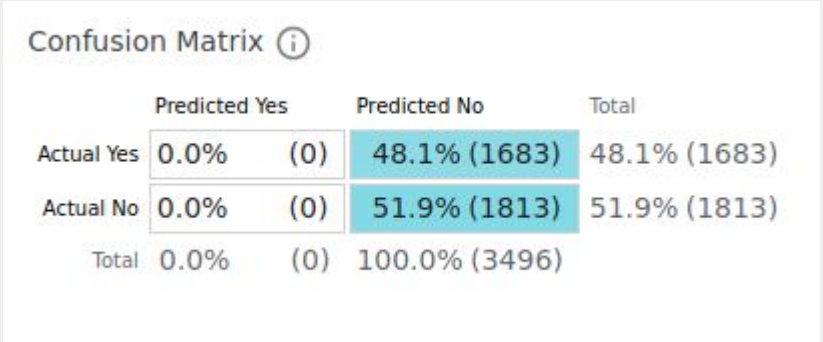
The following 0 to 4 rating scale can be used to rate the severity of usability problems:

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/

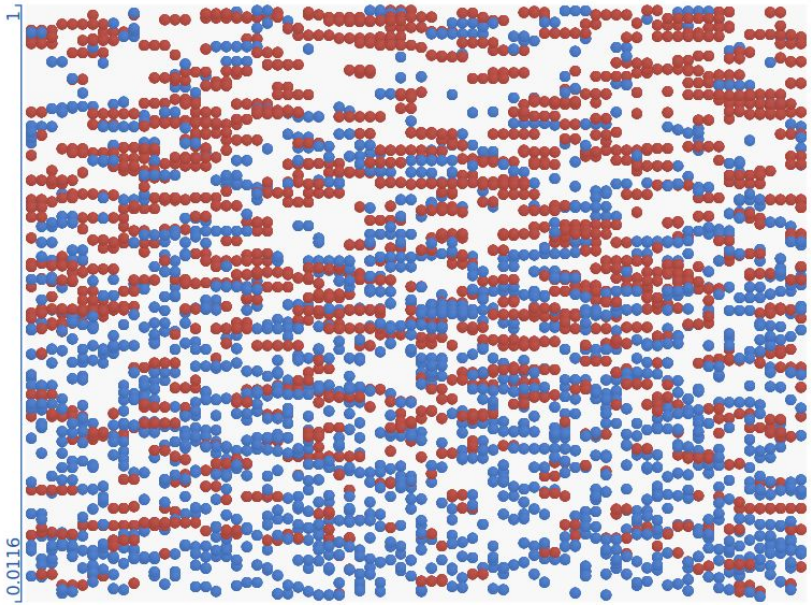
Issue 1: Performance Color Coding

Heuristic:
02 Make clear how well the system can do what it can do.
Description:
48.1% got predicted wrong - this should be clearly stated by the Interface
Severity*: 1
Recommendation:
Change the wrong predicted color to red



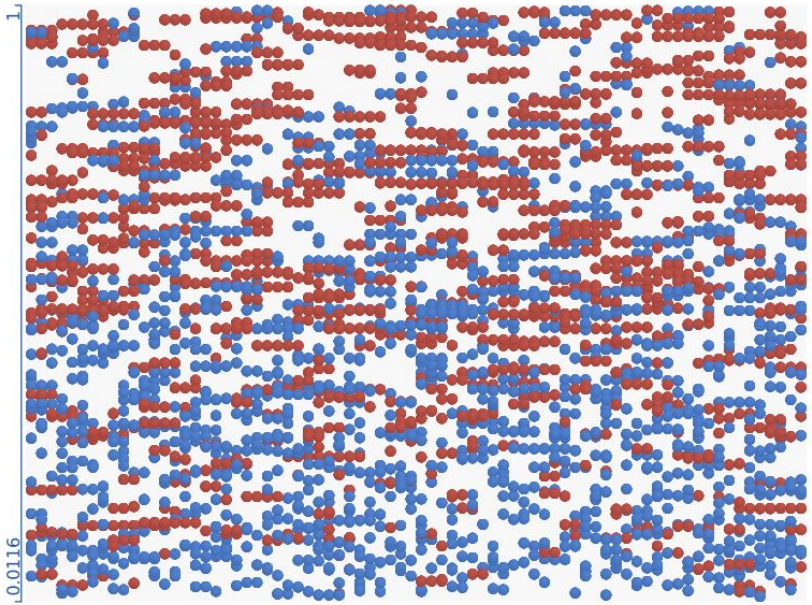
Issue 2: Name of Axis

Heuristic:
04 Show contextually relevant information.
Description:
Blue Axis is not labeled
Severity*: 3
Recommendation:
Show the the names of the axis



Issue 3: Axis ticks

Heuristic:
04 Show contextually relevant information
Description:
No ticks are shown on the blue Axis
Severity*: 1
Recommendation:
Implement an option for showing ticks at relevant positions



Issue 4: Confusion Matrix Formulas

Heuristic:
04 Show contextually relevant information
Description:
Not showing sensitivity, specificity, ppv, npv, … Currently the AI systems performance is based on accuracy but for evaluating performance accuracy is not the single best measurement
Severity*: 4
Recommendation:
Show all relevant formulas for the confusion Matrix

(optional) Screenshot