

## Replies to comments for LDSSA report 2

### Overall impression of comments:

Most sections were fine. In general, main focus is on trimming excesses from sections 2.1 and 2.2. However, the idea that we want a 'shorter' section is counterintuitive (2 pages was asked and given, but 2 pages was too much it seems). Also, requests for alterations to figures for readability.

A main problem is that some of these comments lack specificity/clarity - I can't tell what the problems are, so I have no idea if I am solving them or not. Additionally, in some cases addressing one comment in the way I see fit eliminates the need to address another comment (making it look like I did not address something deliberately, when that was not the case). I think the comments for section 2.1 and 2.2 were a mixture of help and hindrance (some were useful, some were 'reviewer #3' comments)

### 2.1

1. **"We should aim for a more focused/shorter sections (2.1-2.2)."** Text has been reworked, and figures/tables have been condensed.
2. **"This can be achieved by focusing here on precision and recall plots (instead of TPR - can move to annex), without mentioning/assessing the requirements. Should still refer to the threshold you decided in your production model (before getting the new results) and discuss a bit the impact of changing the threshold (more in terms of changes in precision/recall than requirement compliance)".** This has been addressed in the following ways:
  - a. The ROC-AUC curve has been replaced by precision and recall curves from the production model (expected values).
  - b. The requirements (minimum precision, discrimination etc) have been removed from this section and will be exclusively discussed in section 2.2
3. **"It could also help to try and make the text shorter (cutting some bits you feel may be redundant or too detailed - or just moving them to annex)."** Hi Rodrigo, Thanks for the feedback. I have a doubt: the structure of report 4 does not include an annex - should I include one anyway or is it better to just trim the excess and leave the report without an annex?. **"you can include an annex :-) nonetheless, there may be some parts of text that would benefit more of some trimming than just annexing."**
4. **"Also, consider discussing one metric/element at a time and directly compare. Breaking the text in Expected / Achieved, puts each pair of values far away of each other and makes it more difficult to compare".** The text has been rearranged and hopefully this problem is solved. Table 1 also shows both the expected and obtained values now. Additionally, Figure 1 contains the expected and actual results (confusion matrices) for the validation data and the text data.
5. **"Personally, I'm not a fan of figures which are side by side with text, as I believe it distracts from reading it and leads to very thin columns of text."** Figure has been rearranged to break the text (change is only made in sections 2.1 and 2.2 as it was only requested here).
6. **"Also, could you improve blurriness?"** I have had issues with the blurriness when I export from google docs to pdf. I will do my best to resolve this, or to improve readability at least, but I'm not sure whether the problem is me, my machine, or

google docs, so I can't promise success while using this template (you can see it also seems to blur the tables). **No worries with this. As long as you let us know of this issue, it becomes automatically a no-issue. Let's just hope you'll never have to do an actual work report in google docs.**

7. **"How many points has the ROC curve? it seems to have only three, due to the sharp angle. If so, add more points."** Considering your other comments and while trying to trim this section down to size, I am going to remove the ROC-AUC curve (and focus on precision/recall) - consequently, the ROC curve will not be in the next version - is this acceptable, or would you like for me to maintain the plot (in which case I can add it in an annex)? **No reply - ROC curve with more data was included in annex (one plot with expected and actual results plotted)**
8. **"personally, I prefer 80% (subjective)".** Text now reads: "...80%..."
9. **"optimal according to which goal?"**. The method for threshold selection was explained in report 1, the reader is now redirected to that for more information. The word "optimal" has also been removed (it was referring to the threshold which provided the best true positive rate (recall rate), and calculated using geometric means on ROC-AUC curve.
10. **"while this is true, and might mean that your optimal model is not really the best there is, this should not be much related to discrepancy"**. This point (point 3 of the summary list) has been removed from the text.

## 2.2

11. **"Be clear in terms of text, figures and their structuring/layout on what is actual results from the production test and what is already tweaking for a next version."** There is no tweaking for a next version - this is not mentioned at any point, not sure how to deal with that
12. **"Should also move the quantifiable results from 2.2 to 2.1, and here in this section stick with the allowed thresholds and the respective compliance (or not compliance)"**. Sorry to bother you again: I'm not fully understanding which results need to be moved to section 2.1 (or why they need to move). To me it makes sense that if, for example, I say that there was a requirement of 50% precision, I say what was the achieved precision (17%), and that's why it was not met. The same with the discrimination results (which you ask me to not put in the previous section) - it again makes sense to me to say that there was a requirement for a maximum 10% discrepancy, but what we got was 37.5%, and so the requirement was not met etc. Can you explain what you meant so that I can try to address it? **in summary, there should not be any metric result that in 2.2 which is not as well in 2.1. 2.2 is focused on evaluating the compliance of the requirements, but the values that lead to this decision come from 2.1 (which is focused in performance) so, e.g. Figure 4 should be in 2.1 (where you can clarify that this performance is conditioned by the requirements which will be reminded and evaluated in 2.2.** Thanks for clearing that up for me (figure 4 is going to disappear anyway) I will maintain the metrics in section 2.1 as asked. I will also modify Table 2 and discussion of discrimination results, but maintain it in the current location - as it is directly linked to the requirements. **Figure and tables modified to make a cleaner separation between performance and requirements (is hitting a requirement a performance metric? Discrimination results also briefly summarized in section 2.1, but main analysis is left in section 2.2. I expect that this meets the concept of having**

everything included in 2.1, but discuss it in the context of requirements in 2.2. A full summary of these results would be extremely repetitive, so there has to be some balance otherwise the sections have too much overlap (they still do, but that is structural and linked to what was being asked for)

13. **“suggestion on the plot to the right: to minimize white space, customize the xlimits of the line plots”**. This figure is gone, the precision and recall curve data is included with the earlier Figure (Figure 2B). Figure 2B has had x limits customized
14. **“also, to facilitate plot reading:**
  - a. **make the straight lines thinner (and maybe dashed, up to you)**. Straight lines are thinner and dashed
  - b. **avoid having the plot line and the horizontal line in the same/similar color**. Horizontal line showing 50% precision is purple. Line showing threshold at which 50% precision can be achieved is green. Line showing employed threshold (0.5) is blue.
15. **“Personally, I prefer %. Whatever you choose, you should be consistent throughout the table, text and figures”**. Values for precision, recall, accuracy etc have been changed to % in tables and in text

#### 4.1

1. **“Some data analysis on data drift could have been performed, complementing this section nicely”**