

- WHO checked WHAT, and HOW?
  - Ankit, Ankita, Chavi, Kirti and Mabel checked the README.md in the repository, the original paper that the team was replicating, and all R scripts and Markdown files in the analysis folder.
  - Firstly, we went through the ReadMe file to get an overview of the replication work done by the authors of the repository. Then we went through the original authors' paper to get an understanding of the replicated tables and metrics. Finally, we ran the R scripts and the Markdown files in the analysis folder to generate the results locally and compared them with the results shown in the final Markdown file.
- Do the generated outputs match the ones in the paper in the target group's repo?
  - The generated outputs match exactly like the ones in the replication study done by the target group.
- Are the differences relevant or not?
  - There are no differences between the generated outputs in the codecheck and the ones reported by the target group.
  - However, we would like to discuss a little bit about the differences we notice between the results provided by the target group and the ones in the original paper.
  - The replicated F1 scores for the 'hateful' and 'not offensive' classifications are higher than the ones in the original paper, and the F1 scores for the 'offensive, not hateful' and 'overall' classifications are lower than the ones in the original paper. Specifically, there is a 48.8% increase in F1 score for the 'Hateful' classification, a 4.05% increase in F1 score for 'Not offensive' classification, 21.7% decrease in F1 score for 'Offensive, not hateful' classification, and 16.79% decrease in overall F1 score.
  - These differences might be relevant because the authors were using a very limited subset of the total tweets for their analysis, so the model might underfit the data and produce less representative results.
- Are used pieces of software and data properly CITED and with suitable LICENSES?
  - The data, package versions, and dependencies are properly cited in the README. The licenses are not included in the repo.
- Are open formats (text-based) included?
  - Yes, they are. Markdown files and CSV files are included.
- Is data and software FAIR?
  - The data might not be fair, because although the randomly sampled data has the same distribution as the original data, it might possess fewer nuances than the original ones. The authors could have used multiple samples of

3000 randomly selected tweets, or analyzed the original data by chunks under limited computational power.

- The software is fair because they used reasonable libraries in R.