**STANDARDFORM OUTPUT OR EXPORT[[1]](#footnote-1)**

|  |  |
| --- | --- |
| Name *[name user]* | Emily Cantrell |
| Telephone number *[telephone number where the user can be reached when there are questions about this output]* | +33 6 33 02 61 92 |
| Institution *[name institution]* | Rijksuniversiteit Groningen |
| Date *[date on which the output is produced]* | 2024-11-01 |
| Project *[contractnumber + name research]* | 9469 - Predicting Fertility Data Challenge (PreFer) |
| This output consists entirely of documentation *[syntax or text file without research results] If Yes, then you don’t need to fill in the rest of the form* | No |
| This output is an adjusted version of a stopped output. If Yes, then please fill in the date of the stopped output. | No |
| Data *[used datasets]* | Gbapersoontab, gbahuishoudensbus, familienetwerktab, train.csv from PreFer challenge |
| Purpose *[description of the purpose of the analysis in this output]* | The purpose of this analysis is to examine how sample size and the set of predictor variables influence predictive performance for models predicting fertility. |
| Relation with previous output *[description of the relation with previously submitted output]* | This is similar to a previous output we did in which we examined how sample size and feature sets impact predictive performance. This time, we chose different feature sets. |
| Content of the output  *[description of the output, for example: there will be frequencies or crosstabs or regression analysis[[2]](#footnote-2) in this output]* | This output contains:  (1) code (2) a metadata file with notes about the substantive topic area of various variables (3) results: predictive performance metrics in CSV files. Performance metrics are reported on evaluation sets of around 200,000 to 400,000 people. We also include performance metric results from a run where we did hyperparameter tuning on the model. |
| 1 – Does this output contain zeros? | No |
| 2 – Does this output contain values < 10 or is it possible to recalculate values < 10? | No |
| 3 – Is there any group disclosure or dominance in this output? | No |
| 4 – Does this output contain information on 1 institution/company/household/etc? | No |
| 5 – Are there any other points where this output does not meet the [output guidelines](file:///K:\Utilities\Regels_Richtlijnen\Guidelines%20for%20RA%20Output.pdf) ? | No |
| If you have answered “Yes” to at least one of the above 5 questions, please give a short but clear explanation on why there is no risk of disclosure in your opinion. | N/A |

1. If this form is not fully or correctly filled in, the output will not be checked. [↑](#footnote-ref-1)
2. In case the output contains results from models, please clearly mention the underlying number of observations. [↑](#footnote-ref-2)