

SCLPC Peer Review

Data Science Division, Data Ethics Secretariat

Statistics Canada

Introduction

This document provides a peer review of the Spouse or Common-Law Partner in Canada Class (SCLPC) Advanced Analytics project based on the documentation received from the SCLPC team. The core purpose is to inform the team responsible for this prototype of potential ethical and quality concerns. The review has been done by relying on Statistics Canada Responsible Machine Learning Framework. The document entitled “Peer Review for Automated Decision-Making Tools Under Canada’s Directive on Automated Decision-Making”, has also been consulted. Moreover, we have examined the Algorithmic Impact Assessment form as it has been filled by the team responsible for the SCLPC project.

Acknowledgement of best practices implemented by this project

1. An accountability governance model where the Director General of the branch is accountable for the modelling results and their use.
2. Steps taken for model explainability and rules justification reviewed by subject matter experts. Further comments in the recommendations section.
3. A Model Privacy Assessment was completed.
4. A quality assurance plan is in place to monitor the model performance over time.
5. An overview of how advanced analytics is being used in this project has been published in the IRCC webpage for transparency.
6. A gender-based analysis plus (GBA+) was performed for preventing from discrimination.

Major Recommendations

1. A supervised machine learning workflow usually involves training, validation, and test set. There is no mention of a validation set in the documentation. Usually a validation set is used to optimize hyperparameters of the method. Is there any hyperparameters used by the proposed method? If yes, we recommend explaining why and answer any potential leakage concerns that one might have. For example, it would be inadequate to make changes (manually or otherwise) to the final model/decision rules, based on the quality metrics computed with the test set.

2. We recommend addressing the issue of the independence of the observations. Are the training and test set similar to the data that will be obtained in the future? Are there any time/historical dependencies? For example, did the context of the pandemic affect the content of the data in such a way that future data might be very different? If so, statistics computed from the test dataset might be biased.
3. What is the negative impact of wrongly placing a case in the Green Bin? How many individuals could be affected by this mistake? If there is no significant negative impact, then why do we need an interpretable model since the cases that are not placed in the Green Bin will be reviewed by an agent anyway?
4. Bias detection and mitigation
 - 4.1. The document mentions efforts to reduce bias without giving further details. We recommend explaining the type of bias that was considered and examined for this project and how it was mitigated.
 - 4.2. It was not clear if the analytical dataset used for analysis was a sample or a census (whole population).

If a sample,

 - What was the sampling design? Stratification, other? Examples of (good stratification for) representation: time (2018, 2019), treatment centers, programs, other.
 - What was the breakdown of the analytical dataset over 2018 and 2019? Was it 50/50? Other? (We are interested in knowing if there has been a good representativeness over time)
 - In terms of Train/Test, were the proportions 70/30 for both 2018 and 2019?
 - Did we assume that both 2018 and 2019 populations are similar?
 - 4.3. In terms of bias, what is the impact of the post manual adjustments/ tweaks to model rules?
 - 4.4. In page 8, we read that the “model” was built using a dataset of 39,190 applications and then tested on the same dataset to come up with a precision of 99.3%. So here, the performance was not assessed on the TEST set but on the whole data (TRAIN + TEST)? The performance on TRAIN is usually (relatively) high compared to TEST.
5. The document mentions that privacy principles have been taken into considerations (i.e. data minimization, reducing data granularity, de-identification and need-to-know) but it would be good to add some concrete examples that show how this has been implemented.
6. We recommend adding some more quality metrics (confusion matrix for example) and how the challenge of class imbalances has been addressed. Similarly, it would be good to add a few words on the quality of the data and the challenges that it provided (e.g. missing data).
7. To assess the technical part of this work, it would have been nice to have access to a technical report with more details on the data, data processing, modelling, and all the assumptions made.

Minor Recommendation

1. We recommend explaining how this tool will ultimately benefit Canadians in terms that they will understand. By answering the question “Why is it important to reduce the processing time for the applicants?” we can add to the value of the project.