Watchtower Peer Review

Data Science Division, Data Ethics Secretariat

Statistics Canada

Introduction

This document provides a peer review of the Watchtower Risk Identification Prototype. The core purpose is to inform the team responsible for this prototype of potential ethical or quality concerns. The review has been done by relying on Statistics Canada (2021) Responsible Machine Learning Framework. Bronson and Millar (2020) propose a high level process for peer review under Canada's Directive on Automated Decision-Making which was also consulted. Moreover, we have examined the Algorithmic Impact Assessment form as it has been filled by the team responsible for the Watchtower project.

Major Recommendations

1. We recommend describing, with concrete examples, the kind of decisions/actions that the IRCC officials will be able to make/take with this tool in order for the readers to better understand the benefits and the risks associated with this project. What will the agents do with the information? What can happen to the applicants? Are the applicants in Canada or in a different country? It would be good to define the expression 'adverse information' more rigorously. Is it an euphemism of 'crime' or 'offense'?
2. The algorithm associates "data patterns to adverse information" p.5. We recommend using more standard terminology such as "independent variables to dependent variables" or "features to labels". This will improve the interpretability of the documentation of the project.
3. Based on the information in the document, this looks like a predictive machine learning project where one models the relationships between the variables in the dataset. How can this tool provide "fact-based information that end users (IRCC officers) may find relevant for their adjudication process" p.5 if the officers are not going to classify/make predictions on individuals? Expressions such as "identify only those patterns suggestive of organized risk" p.8 imply predictive labelling. Therefore, we recommend removing expressions such as "the tool is not predictive"p.5.
4. If this is a predictive modelling project, then a train/validation/test methodology must be applied. If there is no need for validation, then it should be explained. The metrics that will be used should be described and justified. How do you measure the "utility and timeliness" p.14 ?

5. We recommend a more thorough description of the quality of the data set that will be used by the algorithm. Does the data fail to cover a subpopulation that could be at risk? Are there any missing data? If so, has there been any kind of imputation (without leakage)? Do you use all of the available data to run the algorithm or only the observations that have been previously flagged by officers?
6. We recommend answering the following two questions. Will the information used reflect and exacerbates stereotypes? We understand that the purpose is not to identify "broad subpopulation that tends to have higher adverse rates" but will someone monitor this over time?
7. We recommend describing how the quality of the algorithm will be maintained over time. How will you measure the "utility and timeliness" p.14 over time? Will an individual be investigated over and over again even if the first investigation did not find anything suspicious?
8. It would be good to describe the alternative methods that have been considered, like decision trees, and to explain why they have not been chosen.
9. We recommend that this report should be approved by the Deputy Minister in charge of the program that will use the automated decision system.
10. We recommend the publication of this report and the completed Algorithmic Impact Assessment on the Open Government Portal prior to deploying the system.

Minor Recommendations

1. We recommend explaining how this tool will ultimately benefit Canadians in terms that they will understand. This can help motivate the need for such measures.
2. We recommend indicating that some research has been done (bibliography) on the ethical use of machine learning for decision making with crime data.

Reference

1. Statistics Canada (2021). "Framework for Responsible Marching Learning Processes at Statistics Canada", https://www150.statcan.gc.ca/n1/pub/89-20-0006/892000062021001-eng.htm
2. Bronson and Millar (2020). "Peer Review for Automated Decision-Making Tools Under Canada's Directive on Automated Decision-Making", internal report based on a study with Treasury Board Secretariat and Canadian School of Public Service