Catalogue no. 11-522-x
ISSN: 1709-8211

Proceedings of Statistics Canada Symposium 2021
Adopting Data Science in Official Statistics to Meet Society's Emerging Needs

Responsible use of machine learning at Statistics Canada

by Keven Bosa and Yanick Beaucage

Release date: October 15, 2021

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2021
Adopting Data Science in Official Statistics to Meet Society's Emerging Needs

Responsible use of machine learning at Statistics Canada

Keven Bosa and Yanick Beaucage[1]

Abstract

A framework for the responsible use of machine learning processes has been developed at Statistics Canada. The framework includes guidelines for the responsible use of machine learning and a checklist, which are organized into four themes: respect for people, respect for data, sound methods, and sound application. All four themes work together to ensure the ethical use of both the algorithms and results of machine learning. The framework is anchored in a vision that seeks to create a modern workplace and provide direction and support to those who use machine learning techniques. It applies to all statistical programs and projects conducted by Statistics Canada that use machine learning algorithms. This includes supervised and unsupervised learning algorithms. The framework and associated guidelines will be presented first. The process of reviewing projects that use machine learning, i.e., how the framework is applied to Statistics Canada projects, will then be explained. Finally, future work to improve the framework will be described.

Keywords: Responsible machine learning, explainability, ethics

1. Introduction

More and more data are generated on a daily basis, for example, data from cellphones, satellite imagery, web browsing or optical readers. The profusion of data is paving the way to develop new, more detailed and timely statistics to better serve the population. Like many other national statistical organizations, Statistics Canada has embraced this new reality and is using more and more alternative data sources to improve and update its different statistical programs. Given their volume and the speed at which new data sources are produced, machine learning methods are often required to extract and transform the valuable information contained in these sources.

Statistics Canada has conducted many projects using machine learning methods over the past three years. For example, data scientists used natural language processing to classify comments from census respondents and other surveys and speed up their resolution by automatically distributing them to the appropriate subject-matter experts without having to read them. Unsupervised learning methods were used to cluster the Canadian Coroner and Medical Examiner Database into homogeneous groups to improve understanding of certain events. A supervised learning algorithm was developed to predict crop yield. Projects using neural networks on satellite images are currently underway to optimize the agriculture program by detecting, for example, the presence of greenhouses or identifying the different types of field crops in order to reduce response burden on farmers. An algorithm was also developed to extract financial information from documents. These examples give an idea of different problems where machine learning can facilitate the work of statistical agencies.

There are many benefits to using machine learning: processing large amounts of unstructured data, automating processes, improved coverage and accuracy, and many more. However, it also raises a number of questions, such as:

- Does the process protect data integrity and confidentiality?
- Is the quality of the training data suitable for the desired objective?

[1] Keven Bosa, Statistics Canada, R.H. Coats Building, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6, keven.bosa@statcan.gc.ca ;Yanick Beaucage, R.H. Coats Building, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6, yanick.beaucage@statcan.gc.ca

- Once the algorithm is in place, who is responsible for the results and their effects?

In light of these questions and the increasing use of machine learning methods at Statistics Canada, the Modern Statistical Methods and Data Science Branch recognized the need for a framework to guide the development of machine learning processes and to make those processes responsible.

The Framework for Responsible Machine Learning Processes at Statistics Canada will be presented in section 2. Section 3 will briefly explain the review process put in place to apply this framework. Lastly, this article will conclude with some thoughts and a word on future work in section 4.

This paper is a more detailed version of a paper presented in the July 2021 edition of the Data Science Network for the Federal Public Service newsletter (2021a).

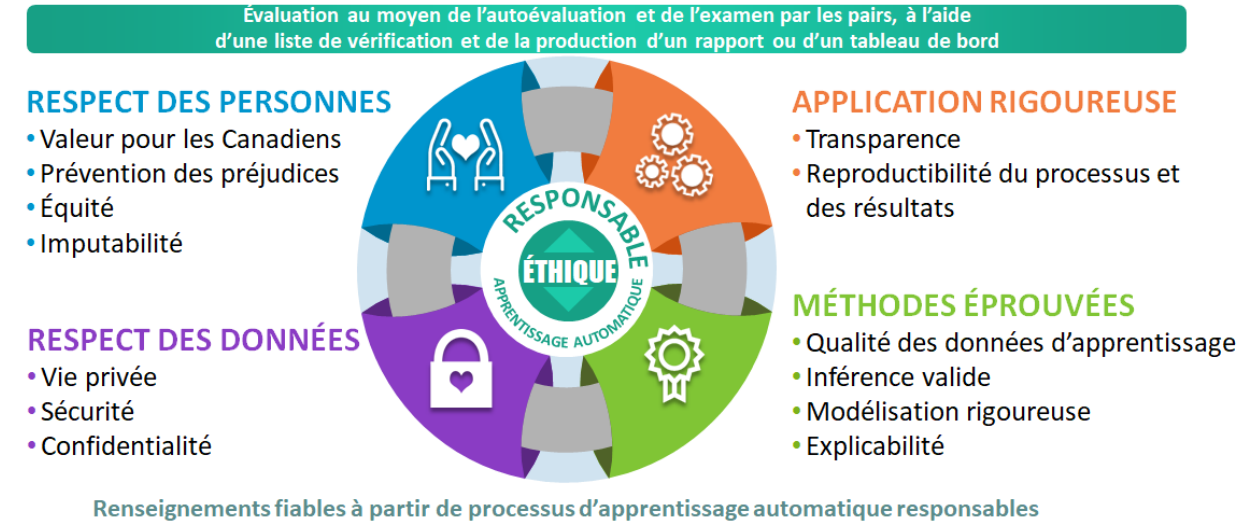
2. Presentation of the framework

Before presenting Statistics Canada's framework, we will give a brief overview of the Treasury Board Secretariat Directive on Automated Decision Making as specified in the Government of Canada document (2019). This directive was the subject of an article in the June 2021 edition of the Data Science Network for the Federal Public Service newsletter (2021b). It states that "the objective of this Directive is to ensure that Automated Decision Systems are deployed in a manner that reduces risks to Canadians and federal institutions, and leads to more efficient, accurate, consistent, and interpretable decisions made pursuant to Canadian law." It also states that the directive "… applies to systems that make or assist in making recommendations or decisions." An algorithmic impact assessment must be completed for projects subject to this directive. The assessment is a self-administered questionnaire in which points are awarded based on the answers provided to each question. The higher the final score, the greater the impact that decisions made by the system could have on the rights, health, well-being or economic interests of individuals or communities, and the more requirements that must be met.

At Statistics Canada, all projects that use machine learning or, more broadly, modelling, are conducted in statistical programs whose purpose is not to make administrative decisions about an individual or an enterprise, but to produce statistics to describe the environment in which Canadians live. As a result, Statistics Canada has not yet had to comply with this directive and assess the impact of these decisions using the Algorithmic Impact Assessment Tool. However, as mentioned at the end of the previous section, Statistics Canada was proactive in adopting this framework to ensure responsible use of machine learning at the agency.

Figure 2-1 provides a good overview of the Framework for Responsible Machine Learning Processes at Statistics Canada: It is organized into four themes: respect for people, respect for data, sound application and sound methods. All four themes work together to ensure the ethical use of both the algorithms and results of machine learning. Each theme is defined by a few characteristics, listed under each one in Figure 2-1. Each characteristic is then defined using one or more guidelines. These guidelines apply to all of Statistics Canada's statistical programs and projects that use machine learning algorithms, particularly those put into production. This includes supervised and unsupervised learning algorithms. Readers should refer to the Framework for the Responsible Machine Learning Processes at Statistics Canada (2021) for further information and the guidelines associated with each characteristic.

Figure 2-1
Framework for Responsible Machine Learning Processes at Statistics Canada



At Statistics Canada, we aim to make efficient use of government resources while producing information that helps Canadians better understand their country. The **respect for people** theme includes four characteristics: value for Canadians, prevention of harm, fairness and accountability.

1. The concept of **value for Canadians** in the context of machine learning means that its use must create added value, either in the products themselves or through greater efficiency in the production process.

2. **Prevention of harm** requires an awareness of potential hazards and meaningful dialogue with stakeholders, spokespersons and advocates before implementing a machine learning project.

3. **Fairness** implies that the principle of proportionality between the methods and objectives is respected, and that a balance is struck between competing interests and objectives. Fairness ensures that individuals and groups are free from unfair bias, discrimination and stigmatization.

4. **Accountability** is the legal and ethical obligation of an individual or organization to be responsible for their work and to disclose the results in a transparent manner. Algorithms are not accountable; someone is accountable for the algorithms.

Statistics Canada takes data seriously. The **respect for data** theme has three characteristics: privacy of the people to whom the data belong, security of information throughout the data lifecycle and confidentiality of identifiable information.

1. **Privacy** is the right to be left alone, to be free from interference, surveillance and intrusions. When acquiring sensitive information, governments have obligations with respect to the collection, use, disclosure and retention of personal information. The term "privacy" generally refers to information about individuals (as defined in Statistics Canada's Policy on Privacy and Confidentiality, 2020).

2. **Security** is the arrangements organizations make to prevent confidential information from being obtained or disclosed inappropriately, based on assessed threats and risks. Security measures also protect the integrity, availability and value of the information assets. This includes both physical safeguards, such as restricted access to areas where the information is stored and used, and security clearances for employees, as well as technological safeguards to prevent unauthorized electronic access (definition from the Policy on Privacy and Confidentiality, 2020).

3. **Confidentiality** refers to protection from disclosure of identifiable information about a person, business or organization. It implies a relationship of "trust" between the supplier of the information and the organization collecting it; this relationship is built on the assurance that the information will not be disclosed without the individual's permission or without due legal authority (definition from Statistics Canada's Policy on Privacy and Confidentiality, 2020).

Sound application refers to implementing, maintaining and documenting machine learning processes in such a way that the results are always reliable and the entire process can be understood and recreated. This theme has two characteristics: transparency and reproducibility of process and results.

1. **Transparency** refers to having a clear justification for why this algorithm and the training data are the most appropriate for the current study. To be transparent, developers should create comprehensive documentation, including making computer code available to others, without compromising confidentiality or privacy.

2. **Reproducibility of process** means that there is sufficient documentation and code sharing such that the process can be recreated from scratch. **Reproducibility of results** means that the same results are reliably reproduced when all conditions are controlled. There are no ad hoc or human intervention steps that could alter the results.

Sound methods are those that can be relied on to efficiently and effectively produce the expected results. Statistics Canada typically follows recognized protocols involving consultation with peers and experts, documentation and testing in developing proven methods. This theme has four characteristics: quality of training data, valid inference, rigorous modelling and explainability.

1. In a machine learning context, **quality of training data** is measured by the consistency and accuracy of labelled data. Coverage, meaning that the labelled data and descriptions cover the entire span of cases the algorithm will encounter during production, is also important to reduce the risk of bias or discrimination (fairness), and ensure representativity of the variables, which is important in achieving realistic performance measures.

2. A **valid inference** refers to the ability to extrapolate based on a sample to arrive at correct conclusions with a known precision measure about the target population. In the machine learning context, valid inference means that predictions made on never-before-seen data are reasonably close to their respective true values in a high proportion, or in the case of categorical data, predictions are correct in a high proportion.

3. **Rigorous modelling** in machine learning means ensuring that the algorithms are verified and validated. This will enable users and decision makers to justifiably trust the algorithm in terms of fitness for use, reliability and robustness.

4. An **explainable** model is sufficiently documented. The documentation should clearly explain the relationship between the input data and the results. It should also make it possible to determine what conclusions can be drawn or what further investigations can be supported. In other words, an explainable model is not a black box.

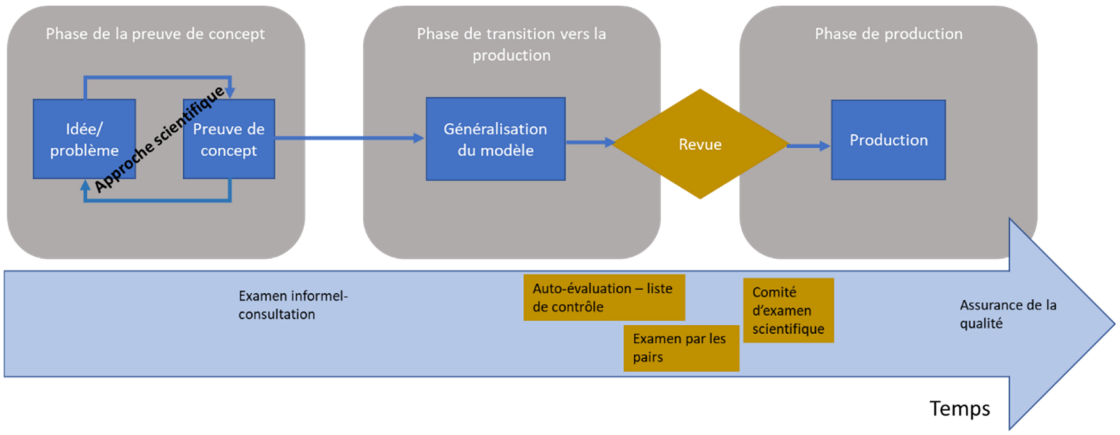
3. Review process

The review process involves implementing the framework at Statistics Canada. The focus is on projects where machine learning methods are used in one or more steps to produce official statistics. Figure 3-1 provides a view of when the review process for a machine learning algorithm should be initiated.

The first box in Figure 3-1 represents the phase during which the proof of concept (or prototype) is developed. Typically, this phase begins with a problem to solve. In this particular context, the proof of concept considers machine learning methods to solve a problem. The scientific approach should be used to develop the prototype. This approach starts with the postulation of hypotheses based on available data. These hypotheses are then subjected to an experiment (or simulation) to obtain results. These are interpreted which leads to a conclusion about the veracity of the assumptions made. Depending on the conclusion, new hypotheses can be considered, which corresponds to the cycle of the scientific approach shown in the grey square to the left of Figure 3-1. When the model developed during the first phase is satisfactory, an informal project review should be done before moving on to the next phase. This review

may take the form of a presentation or consultations with fellow data science experts. If there are no major issues following the informal review, the model where the method will be generalized during the transition to the production phase, which is shown in the second box in the figure. For example, if only part of the data was used during the first phase, the model will be modified to handle the entire data set. The method could also be applied to a larger number of variables during generalization. Towards the end of the transition phase, the formal review of the method should begin. The process includes three steps: self-assessment using the checklist; peer review; presentation of the project to the Modern Statistical Methods and Data Science Branch's Scientific Review Committee. Further information on these steps will be provided in subsections 3.1 to 3.3. Ultimately, the process will recommend whether or not the application should proceed to the production stage. A quality assurance plan should be part of the production phase to continuously review the performance of the model used. Adjustments such as retraining the model or modifying hyperparameters will be necessary if any deterioration in model quality is detected.

Figure 3-1
Visualization of the review process during the lifecycle of a project



3.1 Self-assessment using the checklist

Figure 3.1-1
Checklist

Méthodes éprouvées - *Inférence valide*

English Français

Ligne directrice no	Questions de la liste de contrôle	Oui	Non	Réponse écrite	Autoévaluation remplie par	Évaluation par les pairs menée par	Commentaires formulés pendant l'évaluation par les pairs
27	Décrivez le protocole de validation ou de diagnostic choisi et les mesures connexes et ce qui les rend appropriés compte tenu des objectifs et des mesures d'évaluation du programme statistique et du rôle de l'AA dans ce programme.						
28	Indiquez les mesures d'évaluation et les cibles et décrivez comment les cibles ont été établies.						
29	A-t-on mis de côté des données d'essai adéquates qui n'ont pas été utilisées dans le processus d'élaboration?	○	○				
30	Si les données sur l'apprentissage proviennent d'un échantillon, a-t-on utilisé des poids de sondage? Pourquoi ou pourquoi pas?	○	○				

The team that developed the model or application using machine learning methods must conduct a self-assessment on the use of these techniques. To do so, the team will have to read the framework (see Statistics Canada (2021)) and answer the questions in the checklist. The checklist is a questionnaire where, in general, each guideline in the framework is restated as one or more questions. Figure 3.1-1 shows a portion of the questionnaire on the *Valid Inference* characteristic from the *Proven Methods* theme. When completed, this questionnaire and documentation on the project and the methods used are sent to the team in charge of conducting the review.

3.2 Peer review

The peer review consists of a project review by Statistics Canada experts who were not involved in developing the proposed methodology. Reviewers from two different teams will be involved. On one hand, the questions and documentation related to the first two themes of the framework, "Respect for Persons" and "Respect for Data" will be reviewed by the Data Ethics Secretariat team. The internal or external ethics committee could also be involved depending on the sensitivity and potential risks associated with the project under review. On the other hand, data science experts will be responsible for assessing the method against the themes of "Rigorous Application" and "Proven Methods." At the end of this evaluation, a report with recommendations from all parties involved in the review will be sent to the project manager. The proposed methodology may be re-evaluated by peers based on the recommendations in the report and the responses provided by the project team.

3.3 Presenting the project to the Modern Statistical Methods and Data Science Branch's Scientific Review Committee

The final step in the review process is the presentation of the project to the Modern Statistical Methods and Data Science Branch's Scientific Review Committee. This presentation to a commitee of experts outlines the methodology used in the machine learning process. The objective of this step is to present the project, the methods used and the review to a panel of experts to obtain their approval of the review's conclusions as well as their comments and recommendations to ensure the responsible use of machine learning. The expertise of this committee lies in the "Rigorous Application" and "Proven Methods" themes of the framework. The role of this committee is to review the methodology and challenge it as necessary, including identifying potential gaps or problems and suggesting improvements or corrections. Ultimately, this committee will recommend (or not) the implementation of the proposed methodology in the context of official statistics production.

4. Conclusion

This paper describes the current procedures that have been put in place regarding the responsible use of machine learning at Statistics Canada. New data sources and machine learning methods are emerging almost every day. To remain relevant, the framework presented in this paper will need to be frequently adapted and revised to address emerging ethical and quality issues. Research is currently underway with the goal of improving the framework and making it even more relevant.

One such research project relates to the interpretability and explainability of the machine learning models used. This area has generated a lot of interest and has recently been the subject of several papers. The idea is to do a literature review in this area to properly define these two concepts and to be aware of the methods used to help explain and interpret complex models. Some interpretation methods will then be used on algorithms developed at Statistics Canada to determine the best approach to take. It would also be important to know if more theoretical research should be done and if new interpretation methods need to be developed in the event that the explainability/interpretability of certain models is deemed unsatisfactory, even after the use of methods to aid interpretation proposed in the literature. Another research project involves studying the ready-to-use machine learning solutions offered by some vendors. These tools enable people with little or no knowledge in this field to use machine learning. They can also help data scientists develop a machine learning algorithm more quickly. One of the objectives of this project is to compare the results of some of these generic vendor solutions with specific, in-house developed algorithms. This way, the benefits and risks associated with these solutions can be assessed and identified to prevent undesirable uses of these tools.

In conclusion, machine learning is increasingly considered and used to improve and modernize Statistics Canada's various statistical programs. This makes it necessary to ensure that these methods are used responsibly to preserve the trust of Canadians and to continue to produce more detailed, timely and good quality statistics. It is for this reason that the framework was developed and a review process was established. Training to inform employees about the responsible use of machine learning at Statistics Canada is currently being developed. This will help promote the framework and encourage its use.

References

Government of Canada (2019), "Directive on Automated Decision-Making" URL: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

Data Science Network for the Federal Public Service (2021a), "Responsible Use of Machine Learning at Statistics Canada," URL: https://www.statcan.gc.ca/en/data-science/network/machine-learning

Data Science Network for the Federal Public Service (2021b), "Responsible Use of Automated Decision Systems in the Federal Government," URL: https://www.statcan.gc.ca/en/data-science/network/automated-systems

Statistics Canada (2020), "Policy on Privacy and Confidentiality," URL: https://icn-rci.statcan.ca/31/31a/31a_002-fra.html

Statistics Canada (2021), "Framework for Responsible Machine Learning Processes at Statistics Canada," available on Statistics Canada's internal network at the following address: https://www150.statcan.gc.ca/n1/pub/89-20-0006/892000062021001-eng.htm