**IAOS Young Statistician Prize
Joint First Place**

**Submission for the 2023 IAOS Prize for Young Statisticians**

**Classifying Respondent Comments from the 2021 Canadian Census of Population using Machine Learning Methods**

Ms. Joanne Yoon

Lead Data Scientist

Statistics Canada

<https://www.statcan.gc.ca>

Ottawa, Canada

**Abstract**

To improve the analysis of respondent comments from the Canadian Census of Population, data scientists at Statistics Canada compared and evaluated traditional machine learning, deep learning and transformer-based techniques. Cross-lingual Language Model-Robustly Optimized Bidirectional Encoder Representations from Transformers (XLM-R), a cross-lingual language model, fine-tuned on census respondent comments yield the best result of 89.91% F1 score overall despite language and class imbalances. Following the evaluation, the fine-tuned model was implemented successfully to objectively categorize comments from the 2021 Census of Population, with high accuracy. As a result, feedback from respondents was directed to the appropriate subject matter analysts in near real-time, for them to analyze post-collection.

1. Introduction
2. Methods
3. Data
4. Results
5. Applying fine-tuned XLM-R model to classify Census 2021 respondent comments during collection
6. Conclusion
7. References