Modern Solutions at Statistics Canada

Presentation at the "Conferência Nacional dos Agentes Produtores e Usarios de Dados"

December 3rd, 2025

Statistics Canada / Statistique Canada

Canada

Current Challenges

Relevance, Trust and Resources

Relevance

- **Coverage**
  - Emerging Data Needs (Novelty, Priority, Granularity)
  - Persisting Data Needs
- **Timeliness**
  - Near Real Time & Frequent Data Needs
- **Quality**: "Fit for Use"
- **Competitive Environment**
  - Private Competitors
  - Comparative Advantage in their 'Niche'

Trust

- Maintaining Trust of Data Providers & Respondents
  - Confidentiality
  - Privacy
  - Security
- Maintaining Trust of Data Users
  - Be "The Reference"

Resources

- Meeting Increasing Data Needs with Similar or Fewer Resources

- Financial Resources

  - Cost Savings while Operating Cost Increases

- Human Resources

  - Expertise Profiles: Subject Matter Analysts, Writers, Programmers, Managers

- Technical Resources

  - Stability and Adaptability

Solutions

Solutions with Opportunities and Risks

Solutions, Opportunities & Risks

- Safeguard Confidentiality, Privacy & Security
    - Uphold Trust of Data Providers
    - Limits Granularity
- Competitive Advantage in Neutrality, Rigor & Transparency
    - Ensures Relevance Even if Not Most Timely, Frequent, Detailed
    - Rest on Reputation and Independence
- Use of Alternate Data Sources: Fiscal, Custom, Scanner
    - Increases Coverage & Reduces Costs
    - Vulnerable to Change & Availability

Solutions, Opportunities & Risks (Cont'd)

- Standardize Tools, Concepts & Methods
  - Generates Efficiencies
  - Less Adaptative, Control Loss

- New Techniques: Artificial Intelligence
  - Generates Efficiencies, Increase Capacity, Enhances Skills
  - Ethical and Trust Considerations, Marginal Gains Can be Low (ROI)
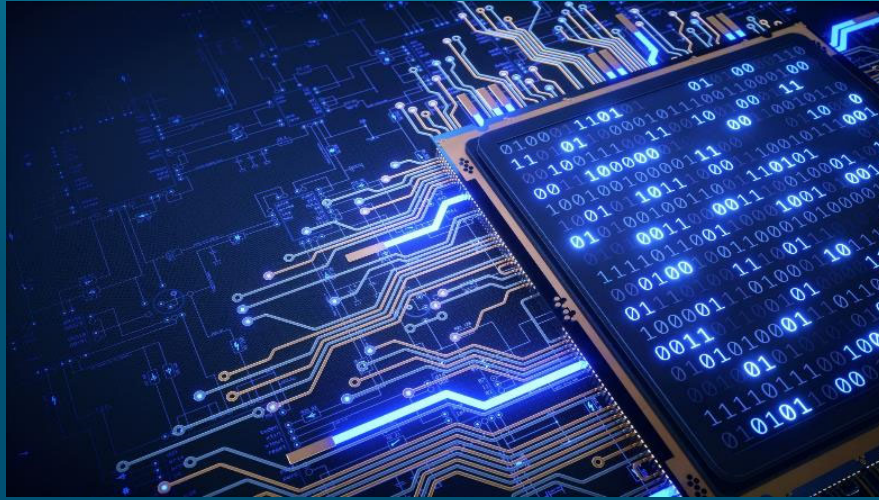
Artificial Intelligence

Recent Uses and Applications at Statistics Canada

Statistics Canada AI Priorities

✓ **Productivity enhancement** - Automate to improve efficiency

✓ **Classification** - Code ingested data into international classifications

✓ **Program conversion** - Convert code (SAS) to open-source language (R and Python)

✓ **Article drafting** - First drafts of descriptive analytical publications, translation

✓ **Data ingestion** - Facilitate the ingestion of unstructured data (comments, Web)

✓ **Testing** - Automate testing of electronic questionnaires

✓ **Chatbots** - Guide users to answers to frequently asked questions (FAQ)

✓ **Web enhancement** - Improve the searchability & discoverability of StatCan website and products

Writing "The Daily" Articles

Objective	AI to draft *Daily* articles and associated administrative tasks
Data used	Publishable data (e.g. value, share and growth by industry segments)
AI/Automation method	AI to generate then refine a first draft for the monthly *Daily* article. Human review remains essential for accuracy and trust.
Results and value added	Savings on Redaction. Enhances analysis
Timeline	Initial setup of 1-2 weeks for design, coding and validation, followed by parallel testing
Status	In production: The Daily — Building permits, The Daily — Investment in building construction, and expending!
Reproducibility	A streamlined workflow that can be applied to any structured report

Product Classification Coder

Objective	Automate product coding process to improve efficiency and accuracy
Data used	Collected product write-ins Historical coded products North American Product Classification System (NAPCS)
AI/Automation method	AI-driven automation to suggest codes and streamline human review
Results and value added	Manual effort reduction Accuracy improved Faster turnaround and consistent data
Timeline	Pilot underway (May 2025 – January 2026)
Status	In pilot production
Reproducibility	Workflow can be adapted to other classification tasks: industry, occupation, etc.

News Analysis Tool

Objective	Identify major investments projects in Canada and extract and compile relevant information for each project
Data used	Online news articles (26 sources currently)
AI/Automation method	Large Language Model (GPT-4o)
Results and value added	Speed and efficiency in research Consistency (bias, interpretation, human error) Most articles identified are relevant Accuracy by variable ranged from
Timeline	Development completed
Status	Pilot production (April 2025 – January 2026).
Reproducibility	Scalable to other programs (relevant variables) and to varying volumes of documents.

Unit Value Corrections for Merchandise Trade

Objective	Perform error detection and imputation of problematic Unit Value (UV) entries for merchandise import transactions
Data used	International merchandise import
AI/Automation method	Extreme Gradient Boosting (XGBoost) machine learning algorithm with specific models for each product category
Results and value added	Notable improvement in quality of value data Reduction in user inquiries Analysts have more time for enhanced analytical activities
Timeline	Development from April 2019 - October 2023
Status	In production since October 2023
Reproducibility	Required significant up-front effort and investment to identify and assess different methodological options, improve the model, and develop quality and assessment tools

Crop Yield Model

Objective	Estimate crop yield over the growing season in Canada while reducing response burden on farm operators
Data used	Weekly coarse resolution satellite data (250m and 1km) Agroclimatic data, provincial crop insurance data StatCan's historical data
AI/Automation method	Robust regression was used to generate crop yield predictions. Data processing was automated and streamlined.
Results and value added	Estimates were equal or better than crop survey estimates Response burden on farm operators reduced to zero Modelled estimates available earlier Generated significant savings
Timeline	Six months for development and testing Two to three weeks to run the models
Status	In production and released via the Daily and CODR table.
Reproducibility	Can be reproduced if the input predictors are available for the region(s) of interest.

AI Lessons Learned

✓ Must Align With Existing Directives and Policies

✓ Governance, Ethics and Peer Review

✓ Early Collaboration and Engagement

✓ To Augment, Not to Replace

✓ Free Up Analysts' Time For Higher-value Tasks

✓ Human-in-the-loop Design

✓ Monitoring and Retraining Model

✓ Transparent Communication

✓ Plan for Scalability as Initial Development can by Onerous

STAY CONNECTED

Eh Sayers podcast

StatsCAN Plus

The Daily

My StatCan

Website

Surveys and statistical programs

Data service centres

Questions? Contact us: infostats@statcan.gc.ca

STATISTICS CANADA

Delivering insight through data for a better Canada