2022: A Machine Learning Odyssey

Emily Wiley, U.S. Census Bureau

AI Day for Federal Statistics

April 30, 2026

Agenda

- A Brief History of ~~Time~~ Economic Programs
- A Briefer History of ~~Time~~ Classification Systems
- The Three-~~Body~~ Model Problem
- The Hitchhiker's Guide to Challenges
- Results, or, I Ran Out of Sci-Fi Puns

United States® Census Bureau

A Brief History of Economic Programs

Economic Census (EC)

- The official measure of the United States' businesses and overall economy
- Around 8 million business establishments, covering most industries and all geographic areas of the US
- Produces estimates of revenue, payroll, employment, and other items

Commodity Flow Survey (CFS)

- Joint effort between the Census Bureau and the Bureau of Transportation Statistics
- Around 165,000 business establishments, covering specific industries
- Produces estimates of domestic freight movement

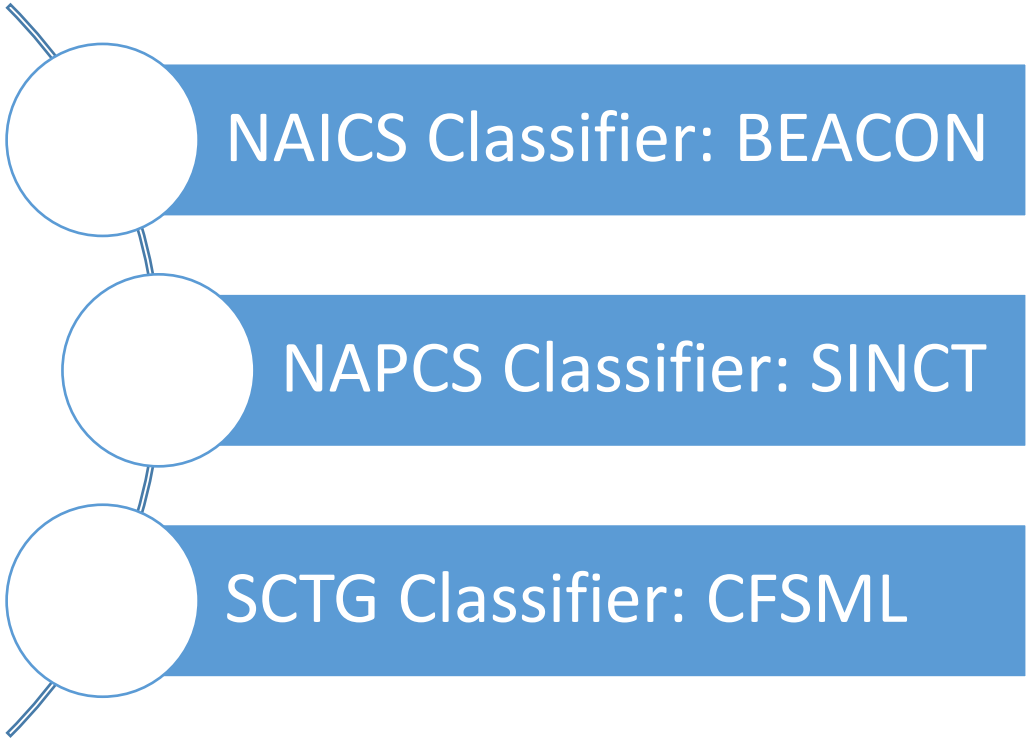
United States® Census Bureau

A Briefer History of Classification Systems

North American Industry Classification System (NAICS)	North American Product Classification System (NAPCS)	Standard Classification of Transported Goods (SCTG)
• Classification system for establishment's primary business activity • Approximately 1,000 NAICS codes • Used in EC and CFS	• Classification system for products and services offered by establishment • Approximately 7,000 NAPCS codes • Used in EC	• Classification system for commodities transported by establishment • Approximately 500 SCTG codes • Used in CFS

United States® Census Bureau

The Three-Model Problem



How are they the same?

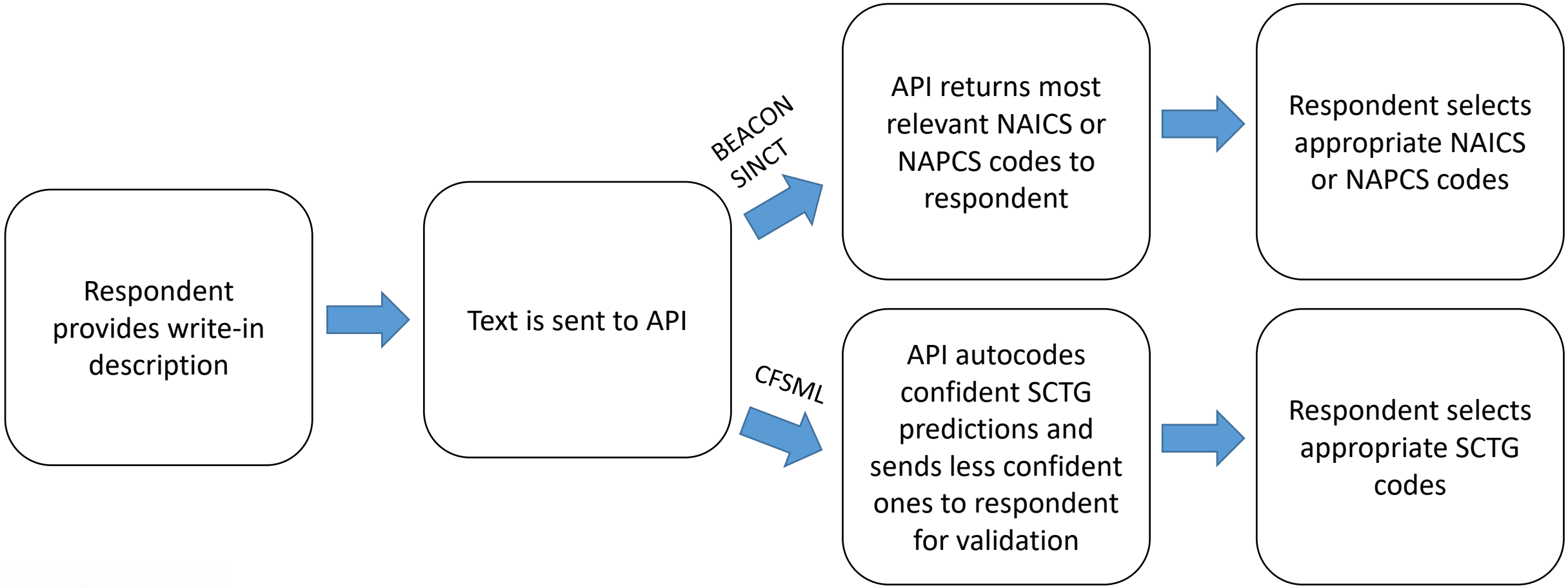
- In-instrument, respondent-facing
- Traditional machine learning
- Fast processing time, calling an API

How are they different?

- Huge differences in training data and classes
- Different types of models
- Respondent interaction is different

United States® Census Bureau

The Three-Model Problem



United States® Census Bureau

The Three-Model Problem: Comparison

	BEACON	SINCT	CFSML
Project Name	Business Establishment Automated Classification of NAICS	Smart Instrument NAPCS Classification Tool	Commodity Flow Survey Machine Learning
Model Type	Ensemble of three information retrieval models	Doc2Vec neural network model	Bag-of-words logistic regression model
Training Data	3.7 million observations	225,000 observations	6.4 million observations
Prediction Classes	Approximately 1,000	Approximately 7,000	Approximately 500

United States® Census Bureau

The Hitchhiker's Guide to Challenges

Speed

- Models needed to run within the instrument, requiring very fast inference
- This constraint limited the complexity and types of models we could deploy

Lack of training data for SINCT

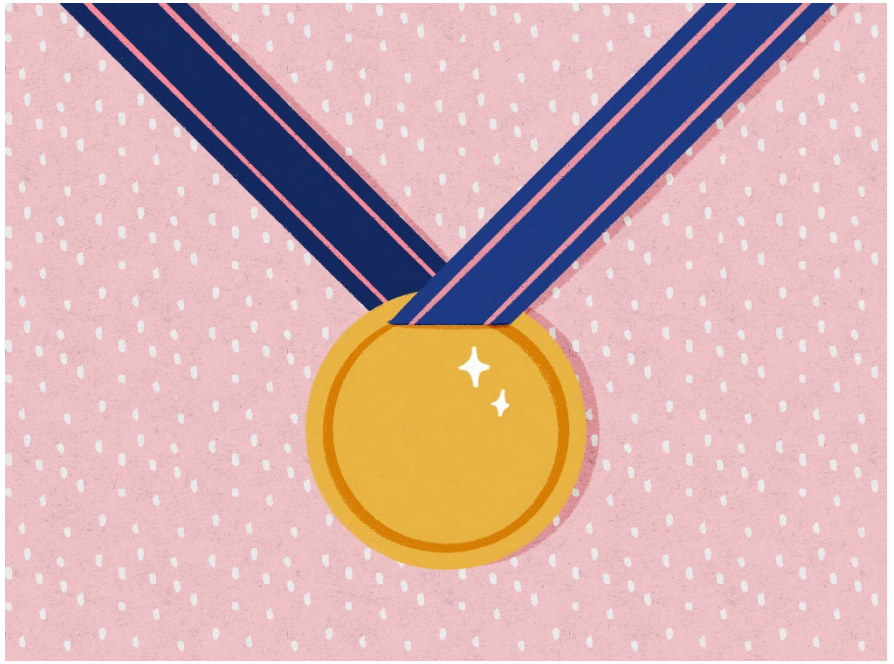
- NAPCS was only introduced in the 2017 Economic Census, resulting in a small labeled dataset
- We expanded training data by mapping a legacy classification system to NAPCS

Disclosure and privacy protection

- Privacy and confidentiality are core principles guiding all Census Bureau work
- Models had to be designed so that no identifying information could be exposed

United States® Census Bureau

Results, or, I Ran out of Sci-Fi Puns



- BEACON:
  - **58% reduction in manual classification**
  - **168,000 write-ins, down from 400,000**
- SINCT:
  - **78% reduction in manual classification**
  - **220,000 write-ins, down from 1,000,000**
- CFSML:
  - **16x increase in shipments collected and classified**
  - **100 million records, up from 6.4 million**

United States® Census Bureau

Thank You!

Emily.L.Wiley@census.gov

United States® Census Bureau

References

- Moscardi, C. (2021, April 13-14). *Modernizing the 2022 Commodity Flow Survey* [Conference presentation]. 2021 Federal Computer Assisted Survey Information Collection Virtual Conference. https://www.census.gov/fedcasic/fc2021/pdf/6a_moscardi.pdf
- Tanner, G. (2024, October 22-24). *Exploring the feasibility of imputation techniques for the Commodity Flow Survey (CFS)* [Conference presentation]. Federal Committee on Statistical Methodology Research and Policy Conference, Washington, DC, United States. https://statspolicy.gov/assets/fcsm/files/docs/2024-conference-docs/I/I6.4_Tanner.pdf
- Wiley, E. (2026, January 20). *Economic Census briefing for Department of Census Taiwan* [Unpublished internal presentation]. U.S. Census Bureau.
- Wiley, E. & Whitehead, D. (2024). *Implementing Interactive Classification Tools in the 2022 Economic Census* [Working paper]. U.S. Census Bureau. https://www2.census.gov/library/working-papers/2024/econ/implementing-interactive-classification-tools-2022-economic-census.pdf

United States® Census Bureau