OEWS Autocoder Overview

Sam Fincher

Branch Chief, Occupational Analytics and Classification Branch

Occupational Employment and Wage Statistics

OEWS Autocoder

- OEWS Data Collection
- Autocoder Background and Overview
- Model History
- Thresholds
- Accuracy Monitoring
- Future Model

OEWS Data Collection

- Occupational Employment and Wage Statistics program
- Federal-State partnership
- Collects and publishes employment and wage data for about 830 occupations at the national, state, and metropolitan area as well as industry level data (NAICS)
- Occupations are coded to the Standard Occupational Classification (SOC) system

OEWS Data Collection

- OEWS fields collected:
  - Job Title (required)
  - Annual or Hourly Wage Rate (required)
  - Description of Duties (optional)
  - Department (optional)
  - Worksite Location (optional)
- Additional establishment level fields available

OEWS Autocoder Background

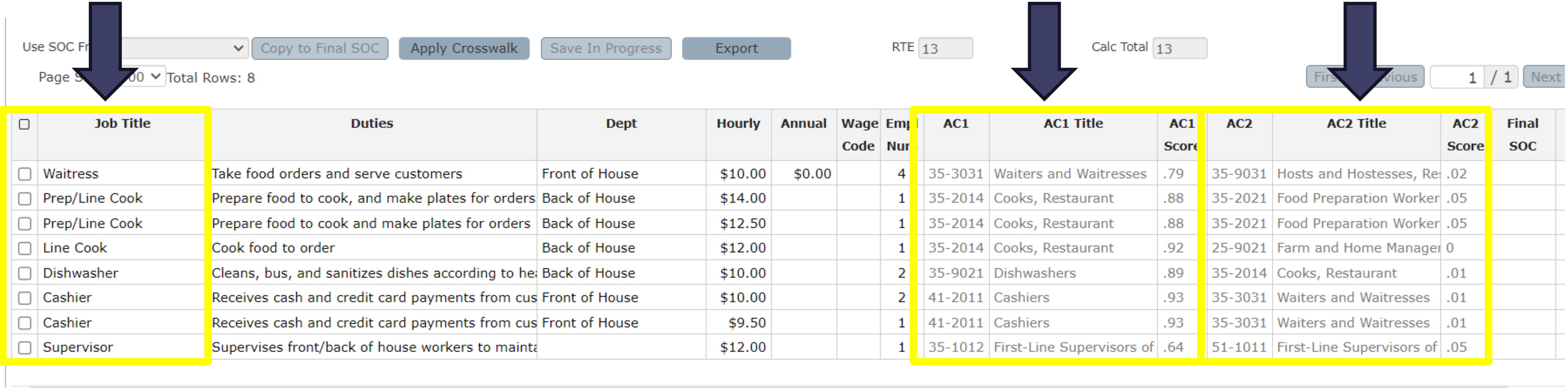
- Assigns Standard Occupational Classification (SOC) codes to respondent data
- Assists states and regions with the labor-intensive process of occupational coding
- Offers two potential SOC codes for job titles meeting the minimum confidence thresholds
- First piloted in 2016 and fully implemented in OEWS production systems in 2021
- Regularly retrained and tested to maintain currency of new job titles and occupations from latest closed panels
- Accuracy continues to improve due to growing training dataset and technological advances

Example of Autocoded Establishment

Reported job title from respondent

Autocode 1 and confidence score

Autocode 2 and confidence score



☐	Job Title	Duties	Dept	Hourly	Annual	Wage Code	Emp Nur	AC1	AC1 Title	AC1 Score	AC2	AC2 Title	AC2 Score	Final SOC
☐	Waitress	Take food orders and serve customers	Front of House	$10.00	$0.00		4	35-3031	Waiters and Waitresses	.79	35-9031	Hosts and Hostesses, Re:	.02	
☐	Prep/Line Cook	Prepare food to cook, and make plates for orders	Back of House	$14.00			1	35-2014	Cooks, Restaurant	.88	35-2021	Food Preparation Worker	.05	
☐	Prep/Line Cook	Prepare food to cook and make plates for orders	Back of House	$12.50			1	35-2014	Cooks, Restaurant	.88	35-2021	Food Preparation Worker	.05	
☐	Line Cook	Cook food to order	Back of House	$12.00			1	35-2014	Cooks, Restaurant	.92	25-9021	Farm and Home Manager	0	
☐	Dishwasher	Cleans, bus, and sanitizes dishes according to he:	Back of House	$10.00			2	35-9021	Dishwashers	.89	35-2014	Cooks, Restaurant	.01	
☐	Cashier	Receives cash and credit card payments from cus	Front of House	$10.00			2	41-2011	Cashiers	.93	35-3031	Waiters and Waitresses	.01	
☐	Cashier	Receives cash and credit card payments from cus	Front of House	$9.50			1	41-2011	Cashiers	.93	35-3031	Waiters and Waitresses	.01	
☐	Supervisor	Supervises front/back of house workers to mainta		$12.00			1	35-1012	First-Line Supervisors of	.64	51-1011	First-Line Supervisors of	.05	

Model History

- Supervised machine-learning, logistic regression model
- Trained on previously coded, labeled job titles
- Three model inputs:
  - Job Title
  - NAICS Code (Industry)
  - EIN (Employer Identification Number)
- Pilot model used approximately 1 million records of training data and 491,000 records of validation data
- Current production model uses 8.1 million records of training data and 4.6 million records of validation data

Autocoder Thresholds

- Thresholds were revamped in 2022 to maximize output to states
  - Autocoder is better at some occupations than others
  - Varying thresholds by SOC Major group
  - Thresholds range from 0.43 to 0.59, and represent the level at which each major group achieves its peak overall accuracy when combined with the accuracy of state coders
- Threshold confidence score must meet one of the following:
  - Autocode1 is at least as high as assigned major group threshold
  - Combined score of Autocode 1 and Autocode 2 achieves 0.65

Autocoder Thresholds

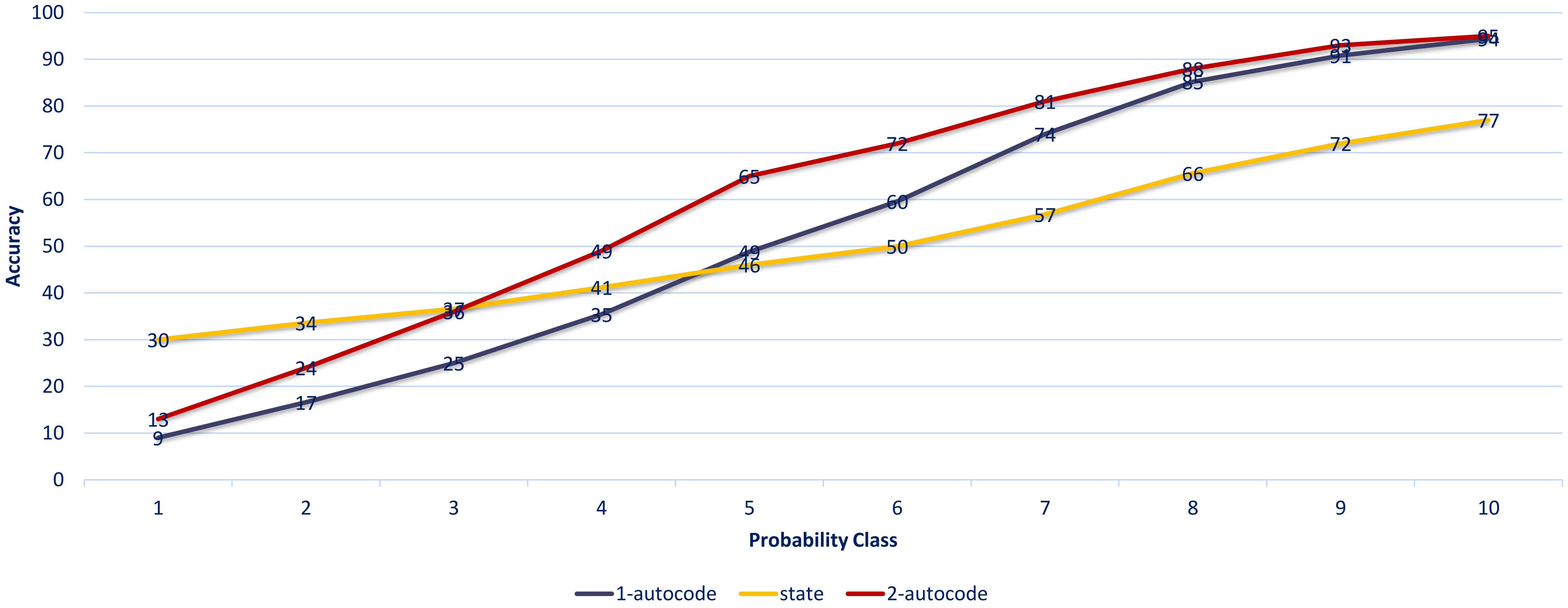
- A lower threshold indicates a higher overall accuracy for a SOC Major Group
- Least confident in Major Group 25, Educational Instruction and Library occupations
- Most confident in Major Group 45, Farming, Fishing, and Forestry occupations

SOC Major Group	Threshold
11 - Management	0.52
13 - Business and Financial Operations	0.55
15 - Computer and Mathematical	0.47
17 - Architecture and Engineering	0.48
19 - Life, Physical, and Social Science	0.50
21 - Community and Social Service	0.57
23 - Legal	0.54
25 - Educational Instruction and Library	0.59
27 - Arts, Design, Entertainment, Sports, and Media	0.49
29 - Healthcare Practitioners and Technical	0.46
31 - Healthcare Support	0.53
33 - Protective Service	0.56
35 - Food Preparation and Serving	0.51
37 - Building and Grounds Cleaning and Maint.	0.52
39 - Personal Care and Service	0.55
41 - Sales and Related	0.53
43 - Office and Administrative Support	0.52
45 - Farming, Fishing, and Forestry	0.43
47 - Construction and Extraction	0.49
49 - Installation, Maintenance, and Repair	0.47
51 - Production	0.55
53 - Transportation and Material Moving	0.48

Accuracy Monitoring

- "Gold-Code" dataset used to separately test the accuracy of the Autocoder and determine thresholds
- The Gold-Code dataset contains job titles that are expertly coded and agreed upon by two state coders without assistance from the autocoder
- This process brings an impartial review of overall coding accuracy of both the autocoder and human coders
- Current Gold-Code file contains 60,000 titles and we are currently collecting an additional 16,000

Probability Class Accuracies



Monitoring Autocode Use

- Autocoder State Report provided to the field each panel
- Tracking percentage of autocodes being used as final production code
- Encourage use of Autocoder, but careful coders are not “blind coding”
- All autocodes need to be reviewed by humans before applying
- November 2023 Panel
  - Out of 3,548,635 job titles, 2,447,686 (69%) above thresholds and provided an autocode
- The final chosen SOC code matches an autocode **87.7%** of the time
  - When provided a code, Autocode 1 matches the final SOC code 78.9% of the time
  - Autocode 2 matches the final SOC code 8.8% of the time

Future Model

- In the process of transitioning to new model
- Convolution Neural Network (CNN)
  - Job titles are vectorized and used to create embedding matrixes
- Early results are promising, with 4% increase in overall accuracy from current production model
- Currently conducting model comparisons
  - New model has higher accuracy in lower probability classes

Contact Information

Sam Fincher

fincher.samuel@bls.gov

Occupational Analytics and Classification

OEWS_OAC@bls.gov