

Consumer Expenditure Surveys Diary Autocoder

THE DIARY SURVEY

ILLUSTRATIVE DIARY

What did you buy or pay for?	Cost without tax	Was the item for:					Name of Store or Website where purchased	
		Child Under 2	Boy 2-15	Girl 2-15	Man 16 & over	Woman 16 & over		
jeans	75 00	1	2	3	4	5	X	Dillard's.com
running shoes	69 00	1	2	3	4	5	X	↓
wallet	29 00	1	2	3	4	5	X	
baseball cap	14 99	1	2	X	3	4	5	Target
bib	3 50	1	X	2	3	4	5	Sweet Dreams boutique
necklace	250 00	1	2	3	4	5	X	Olde Towne jewelry
non-prescription sunglasses	59 00	1	2	3	4	5	X	Walmart.com
child's costume (returned for refund)	15 00	1	X	2	3	4	5	Partysupply.com

- CE collects information on the complete range of consumers' expenditures, income, and demographic characteristics
- Diary survey captures smaller, more frequently-occurring expenditures recorded by respondents over a two-week period
- Four broad categories: clothing, food & drink at home, meals outside the home, all other purchases
- Data are collected and digitized by our partners at Census before being transmitted to BLS

EXAMPLES OF FREE-FORM DESCRIPTIONS

sweatpants
bell-bottoms
corduroys
joggers
cargos
maternity pants
Yoga tights

PANTS Work pants
JEANS Linen trousers
High-waisted flared denim
DRESS SLACKS

All are grouped together in
SLACKS
Item code 410060

(BLS internal item code structure)

OPPORTUNITY

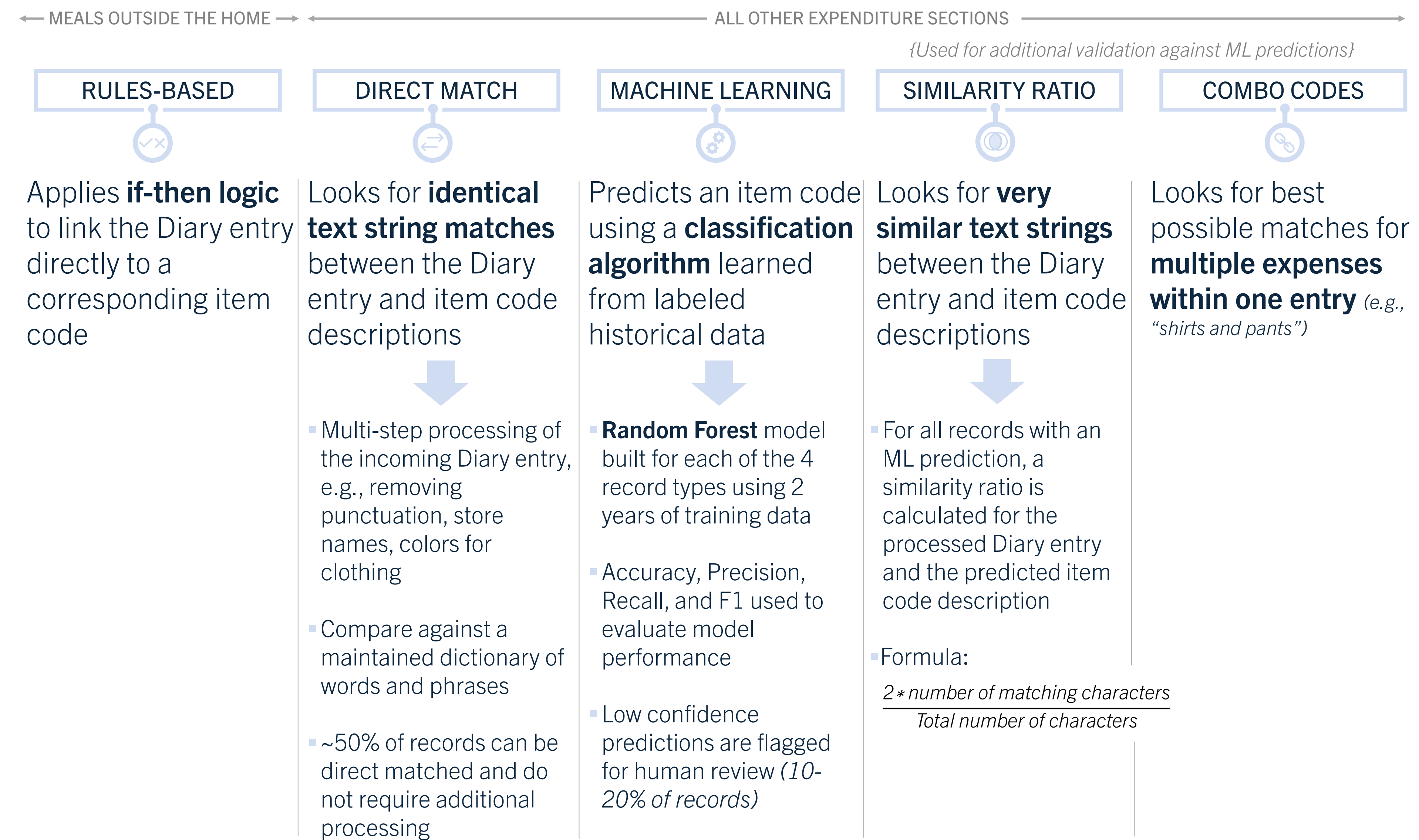
WHAT 30K Diary entries are manually keyed in and coded per month

HOW LONG 8 weeks of manual keying and coding per month of data

ACCURACY 11.8% of records are misclassified despite built-in adjudication

Can we reduce processing time and resourcing costs while maintaining or improving data accuracy?

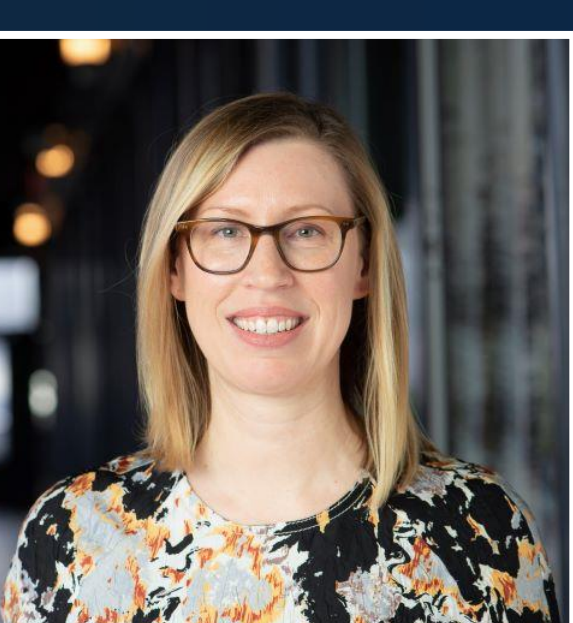
THE DIARY AUTOCODER



Accuracy is the same as manual coding
Estimated 73% reduction in processing time and 13% reduction in costs

LEARNINGS

- Human intervention is an important component of the model building and output review processes
- Model size matters given the infrastructure
- Deployment planning must be thoughtful, involve all stakeholders, and offer generous time frames
- Troubleshooting in production: the ability to react in a timely manner
- Planning for the future: server needs and staff skillsets to support Autocoder



Melissa Pollock
Economist and CE Diary Autocoder Product Owner
Bureau of Labor Statistics
pollock.melissa@bls.gov

Visit us at:
<https://www.bls.gov/cex/>

