Snorkel DryBell: A Case Study in Deploying Weak Supervision at Industrial Scale

Stephen Bach (Brown University); Daniel Rodriguez (Google); Yintao Liu (Google); Chong Luo (Google); Haidong Shao (Google); Cassandra Xia (Google); Souvik Sen (Google); Alex Ratner (Stanford University); Braden Hancock (Stanford University); Houman Alborzi (Google); Rahul Kuchhal (Google); Chris Ré (Stanford University); Rob Malkin (Google);

BROWN

Google

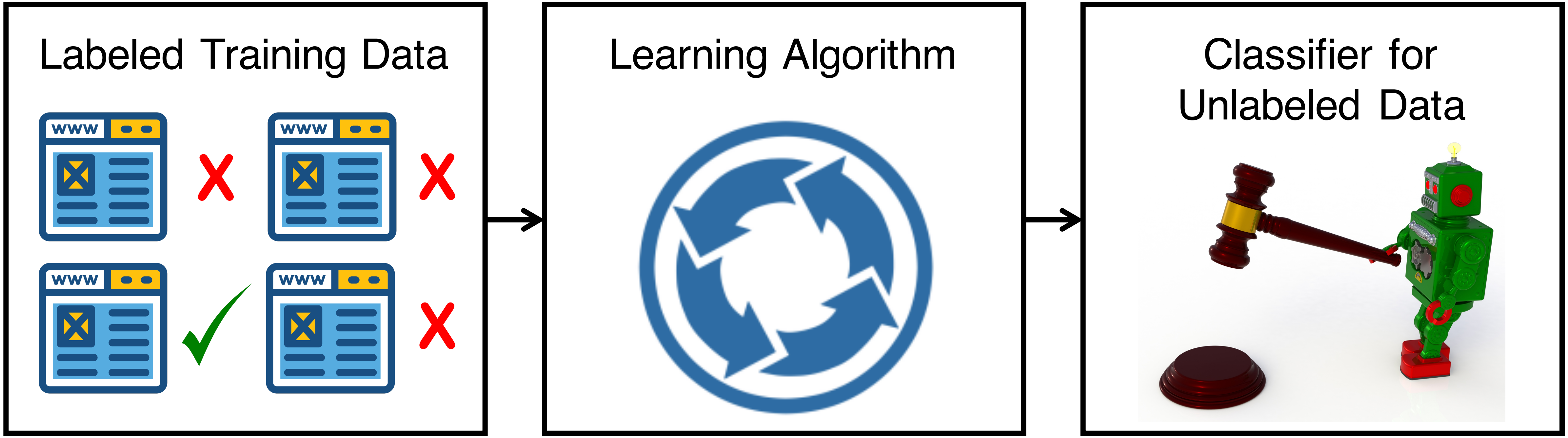
Stanford

This Talk

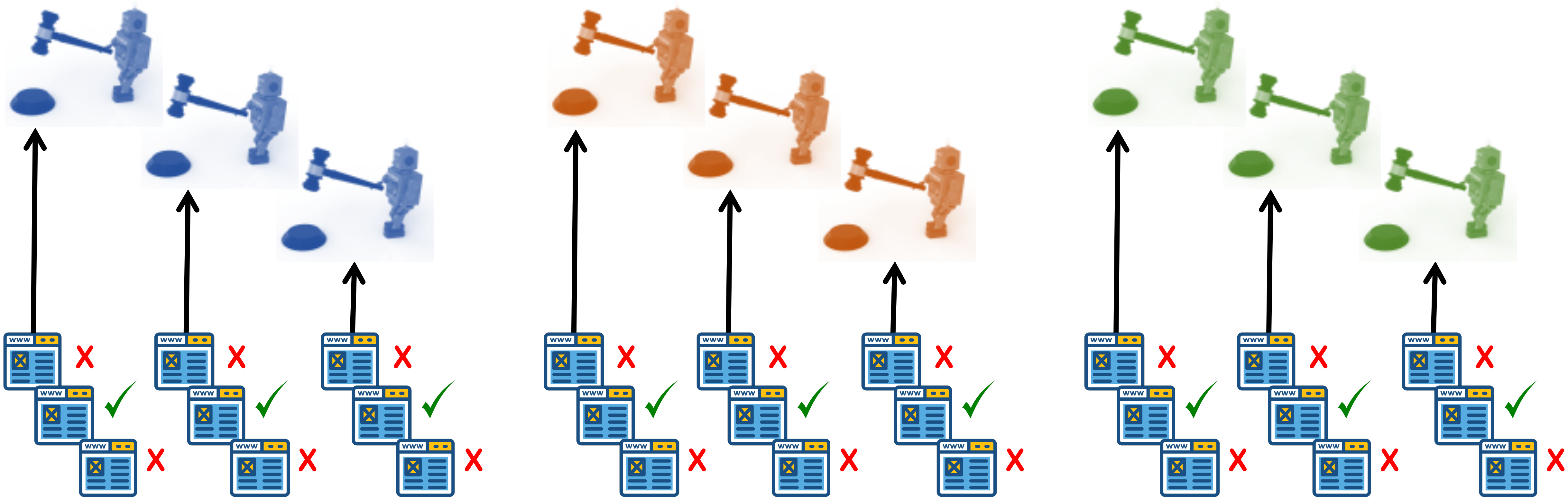
- **Weakly supervised machine learning** seeks to train classifiers **without hand labeled training data**
- What **impact** can it have on **industry** and other **organizations** that use machine learning? What **challenges** arise?
- It can save labeling **tens of thousands of examples** without sacrificing prediction quality!

Training Data is the Bottleneck for Industrial Machine Learning

Supervised Machine Learning



Today's Organizations: Many Classifiers



Weak Supervision with Rules

Open-Source Framework: Snorkel

- Open-source framework to program classifiers by writing **rules that label data**

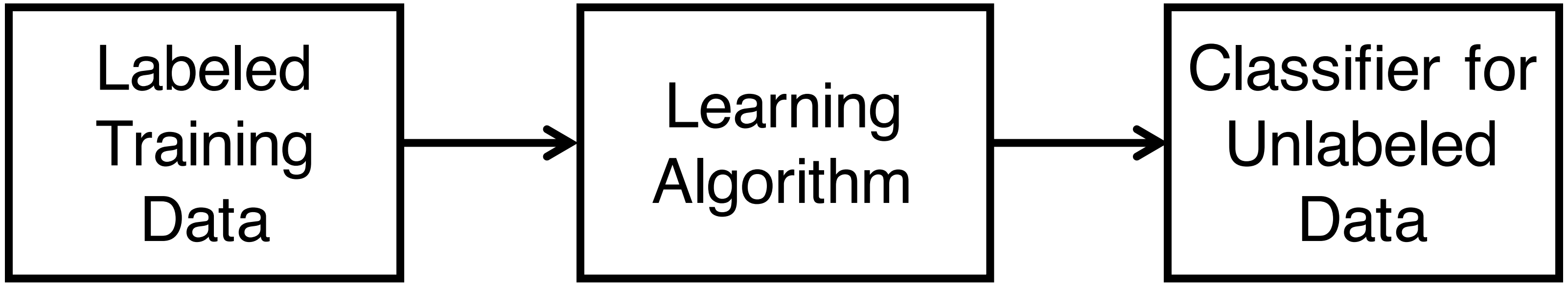
- Results: State-of-the-art performance on benchmark tasks and new applications **without any hand-labeled training data**

snorkel.stanford.edu

Snorkel: Rapid Training Data Creation with Weak Supervision. A. Ratner, S. H. Bach, H. Ehrenberg, J. Fries, S. Wu, C. Re. PVLDB 11(3):269-282, 2017. **Best of VLDB 2018**

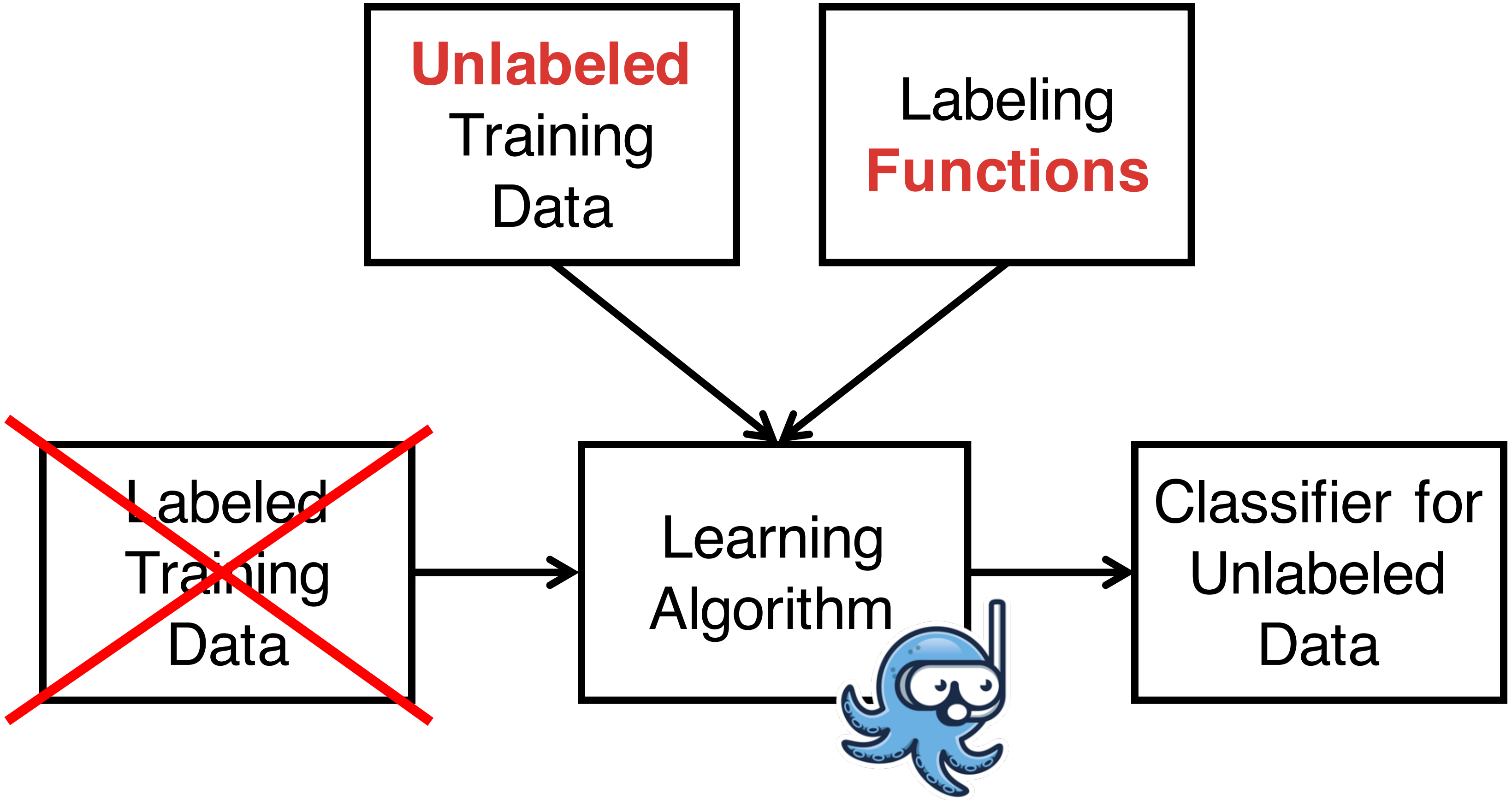
Supervised Machine Learning

Pipeline:



Weakly Supervised Machine Learning

Pipeline:



Example Task: Celebrity News

	True Label
Lori Loughlin's 'Fuller House' Future Dim Due To Elite School Bribery Scandal	✓
HOW TO WATCH TESLA'S MODEL Y REVEAL TONIGHT	✗
Morris Dees, a Co-Founder of the Southern Poverty Law Center, Is Ousted	✗

Example Labeling Function: Keywords

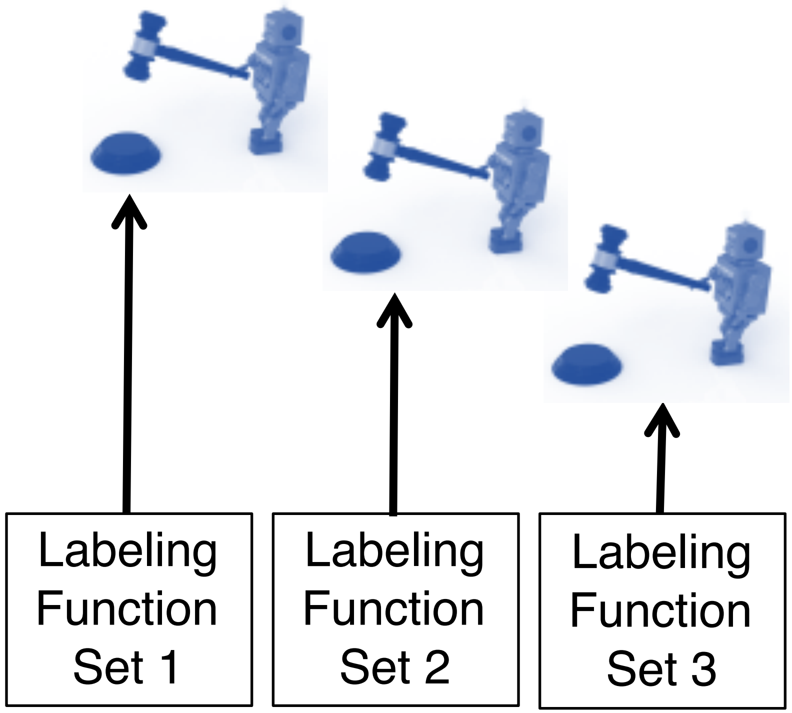
Headline	Vote
Lori Loughlin's 'Fuller House' Future Dim Due To Elite School Bribery Scandal	✓
HOW TO WATCH TESLA'S MODEL Y REVEAL TONIGHT	?
Morris Dees, a Co-Founder of the Southern Poverty Law Center, Is Ousted	✓

In the Industrial Setting…

How Can We:

1. Manage the Proliferation of Supervision Sources?

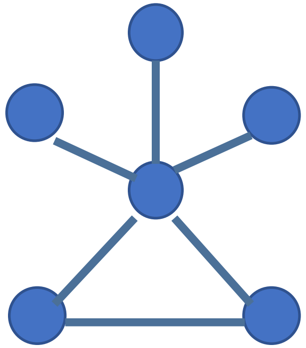
2. Turn the Many Overlapping Sources into an Advantage?



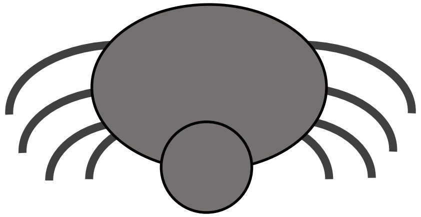
Don’t Start from Scratch!

Knowledge Resources

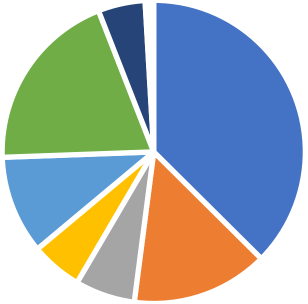
Related Classifiers



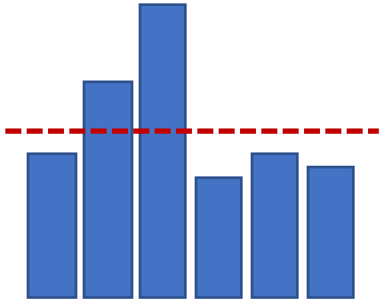
Knowledge Graphs



Web Crawlers



Topic Models

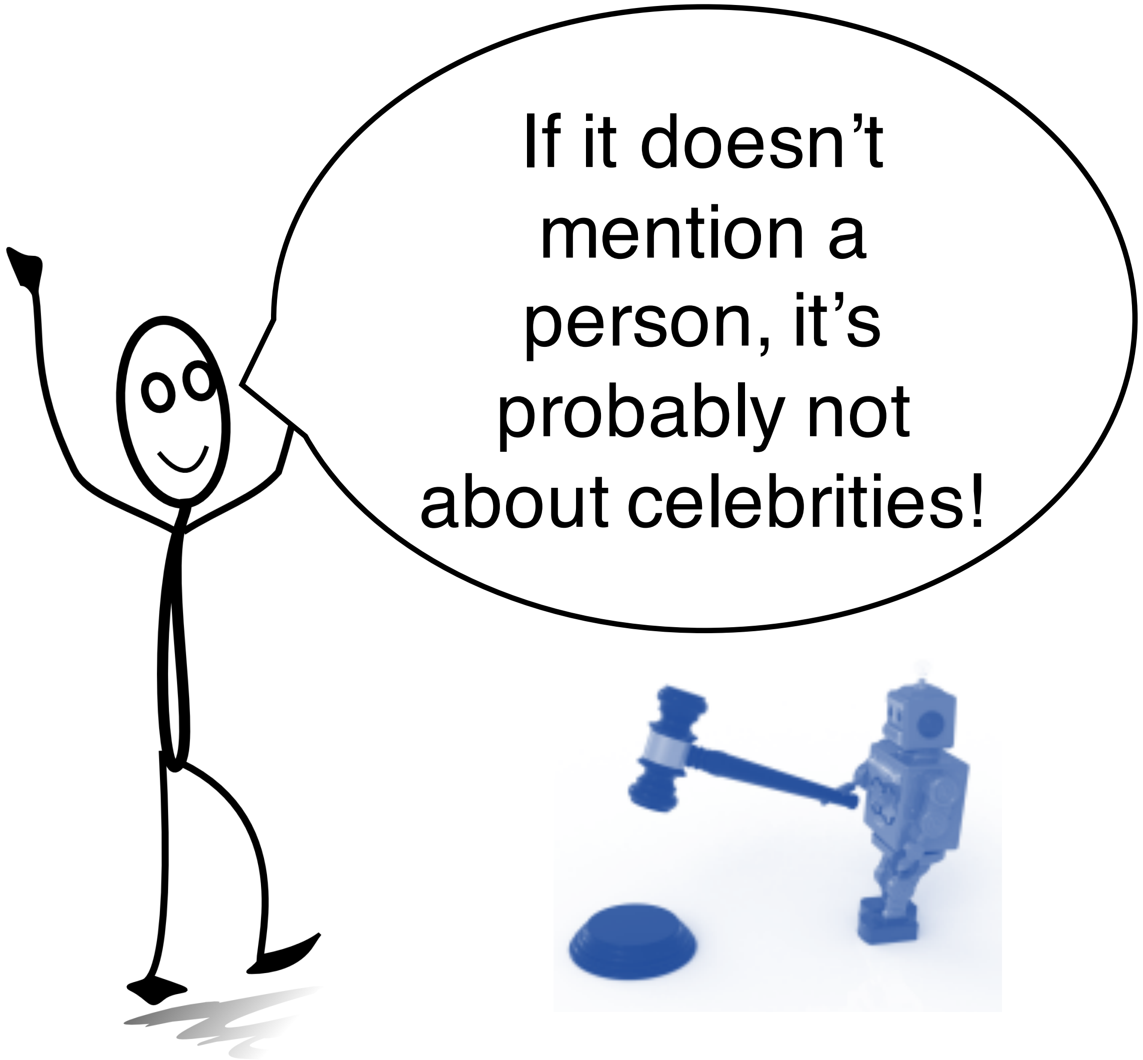


Aggregate Stats

```
If Pattern(data)
Then
data.label = True
```

Rules

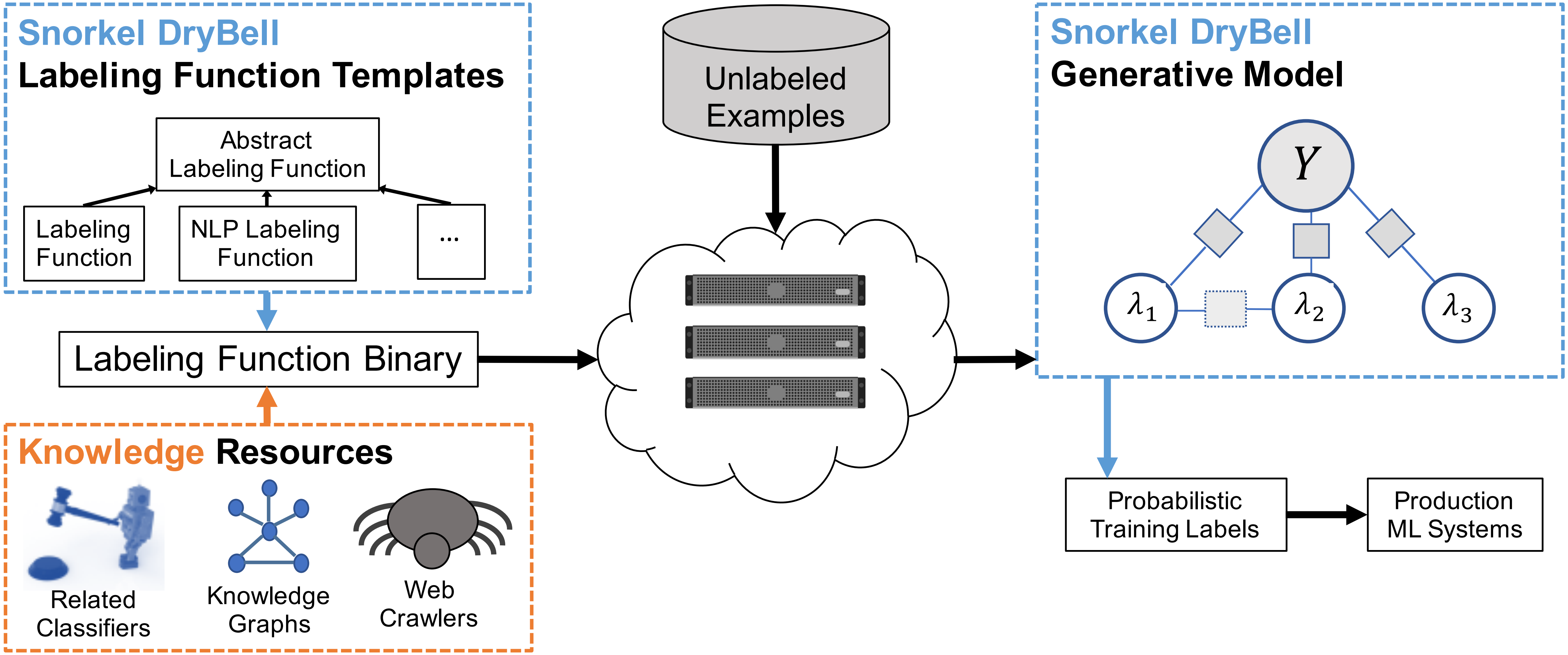
Example: Related Classifier



	Vote
Lori Loughlin's 'Fuller House' Future Dim Due To Elite School Bribery Scandal	?
HOW TO WATCH TESLA'S MODEL Y REVEAL TONIGHT	X
Morris Dees, a Co-Founder of the Southern Poverty Law Center, Is Ousted	?

Snorkel DryBell

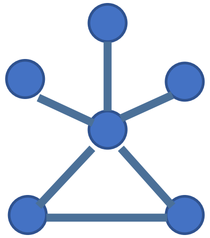
Snorkel DryBell Architecture



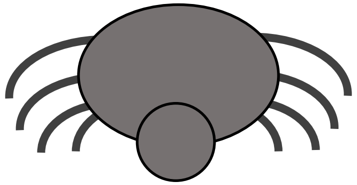
Resources Come in Diverse Forms

- Related classifiers need their own servers

- Knowledge Graph has REST API



- Web crawlers maintained by separate team



Snorkel DryBell Provides Templates

Example: NLP Labeling Function

```
string GetText(const Example& x) {
    return StrCat(x.title, " ", x.body);
}

LFVote GetValue(const Example& x,
                const NLPResult& nlp) {
    if (nlp.entities.people.size() == 0) {
        return NEGATIVE;
    }
    else { return ABSTAIN; }
}

int main(int argc, char *argv[]) {
    Init(argc, argv);
    NLPLabelingFunction<&GetText, &GetValue> lf;
    lf.Run();
}
```

Defines text to analyze

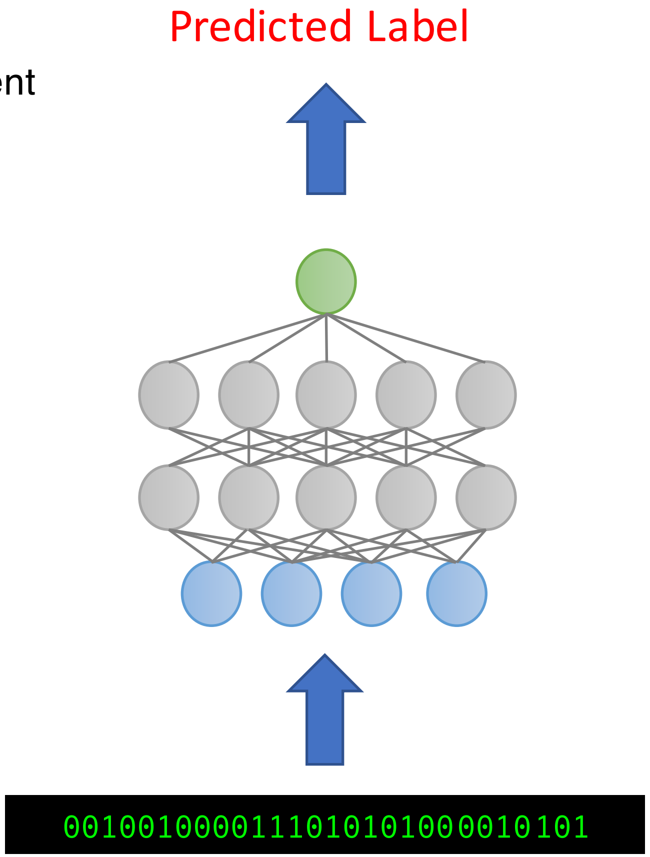
"If the text doesn't mention any people, vote negative"

Launches MapReduce pipeline, starts NLP classifier server on each worker, and saves the results

Resources are Often Not Servable

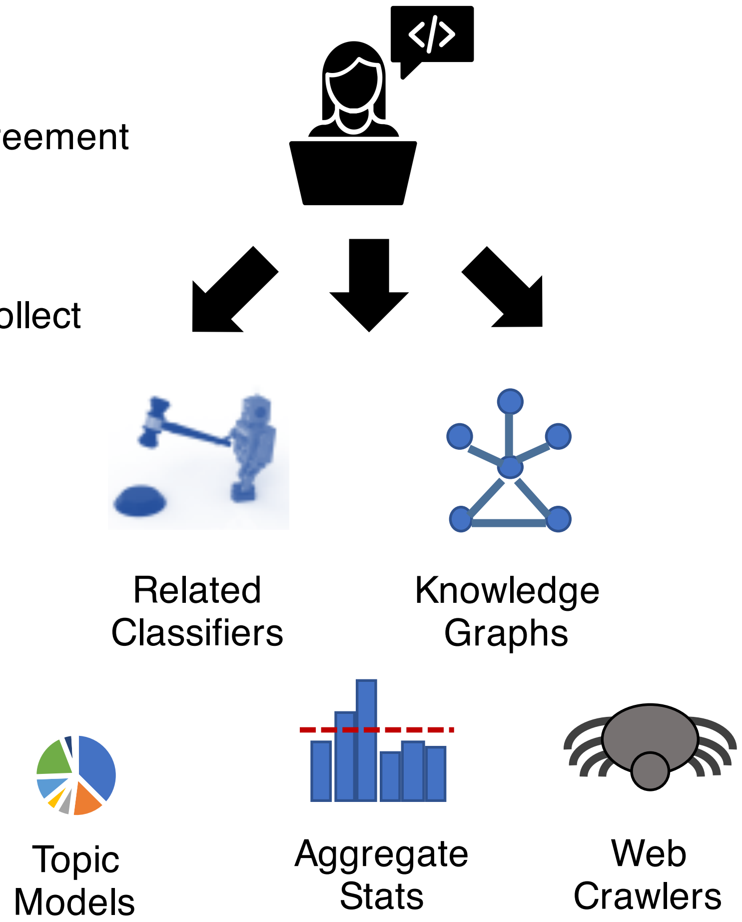
Servable

- ✓ Service-Level Agreement
- ✓ Fixed Model
- ✓ Fixed-Size Input

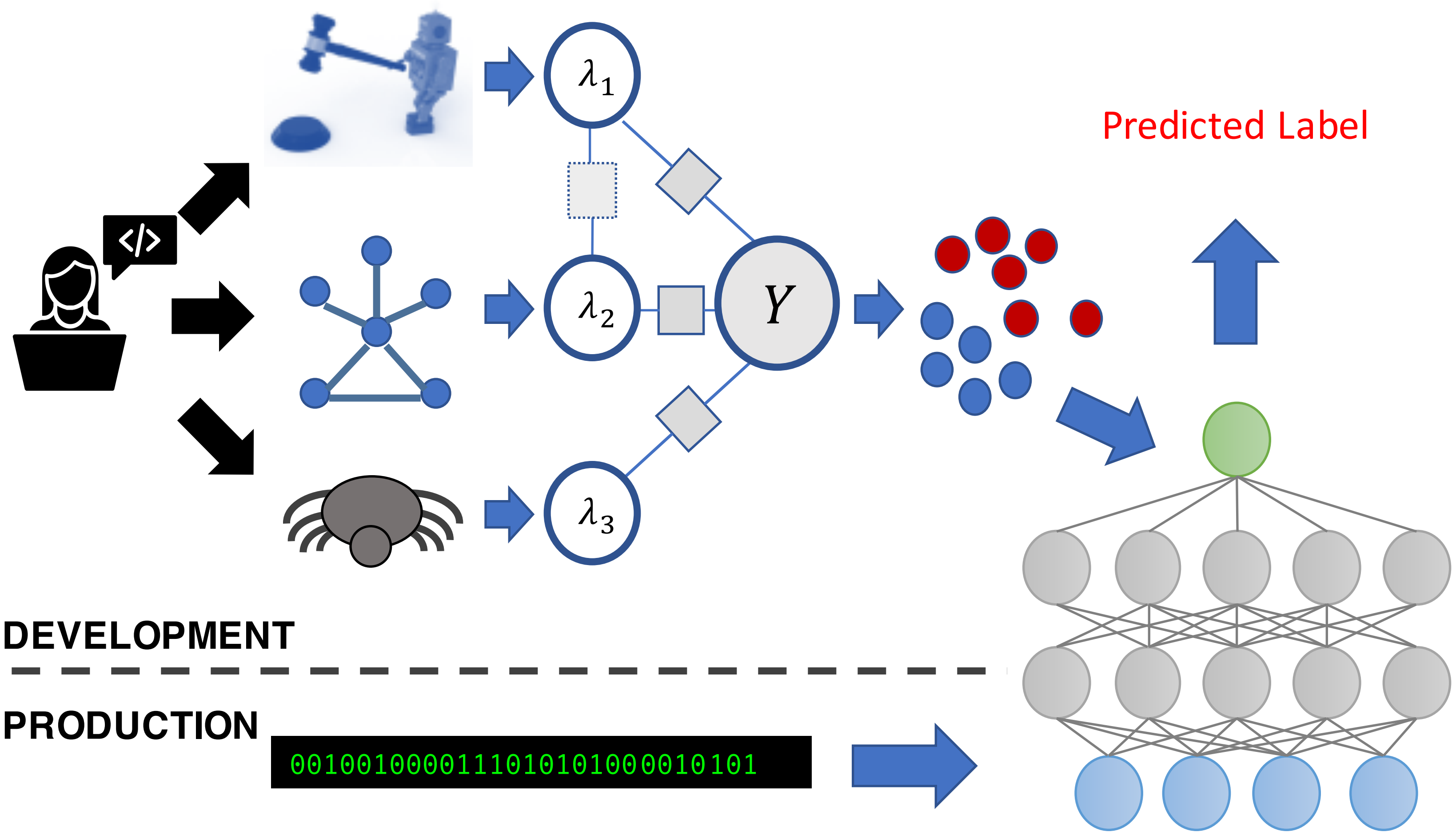


Not Servable

- ✗ No Service-Level Agreement
- ✗ Input Varies in Size
- ✗ Input Expensive to Collect



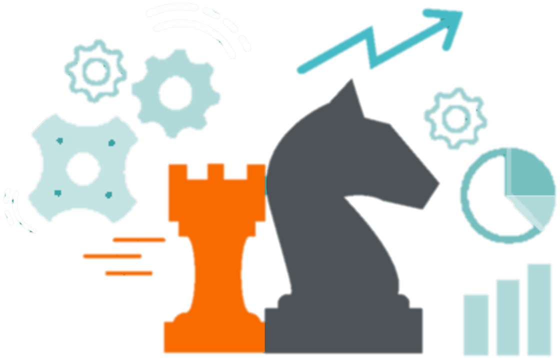
Knowledge Transfers to Servable Models



Experimental Study

Case Studies at Google

- Collaborated with an engineering team responsible for 100+ classifiers in production



- Looked at two recent instances where strategic decisions necessitated new classifiers

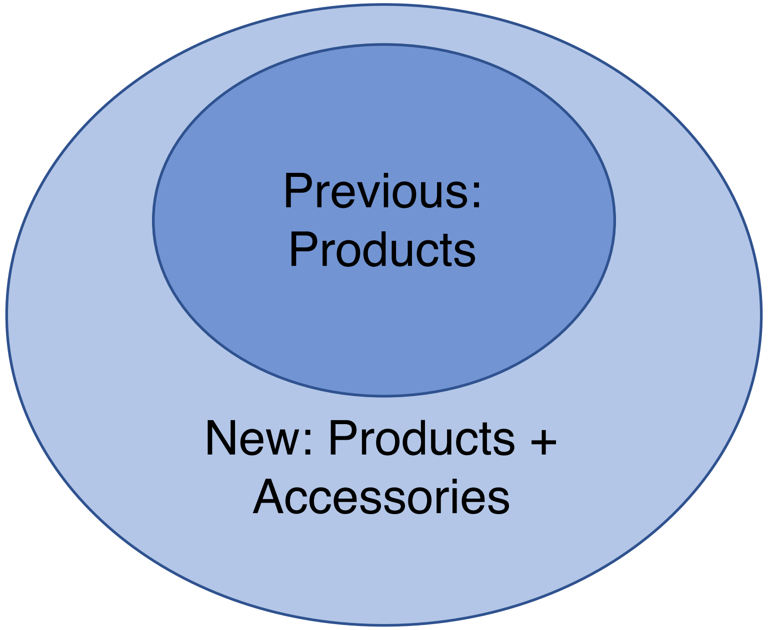
- Due to sensitive nature of applications, we describe at a high-level and report relative scores

Case #1: Product Classification

- Existing classifier used to detect products in a certain category of interest

- Goal: expand label to include accessories

- Instant depreciation of investment in labels!

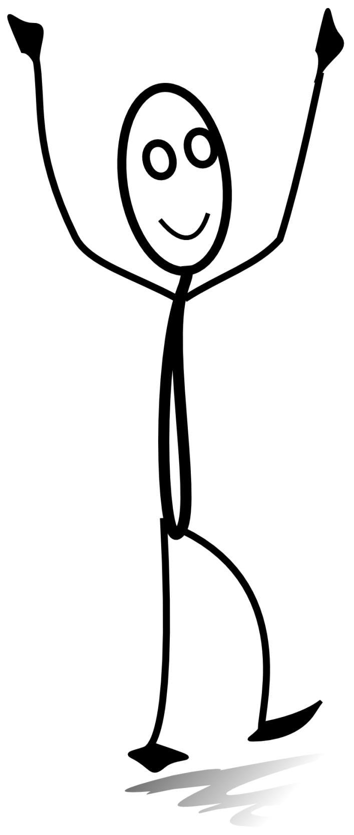


Case #2: Topic Classification

- Emerging topic of interest in Google content

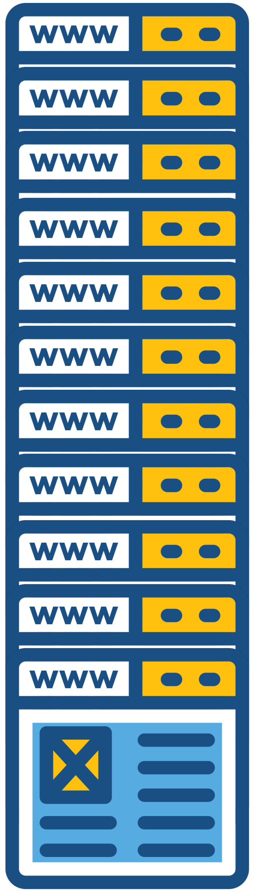
- Goal: develop new classifier to identify topic

- Default procedure is to collect hundreds of thousands of labels for new topic!



Setup

Since these are production tasks, large labeled data sets were available



Hundreds of thousands to millions of examples for training data, which were **treated as unlabeled by Snorkel DryBell**

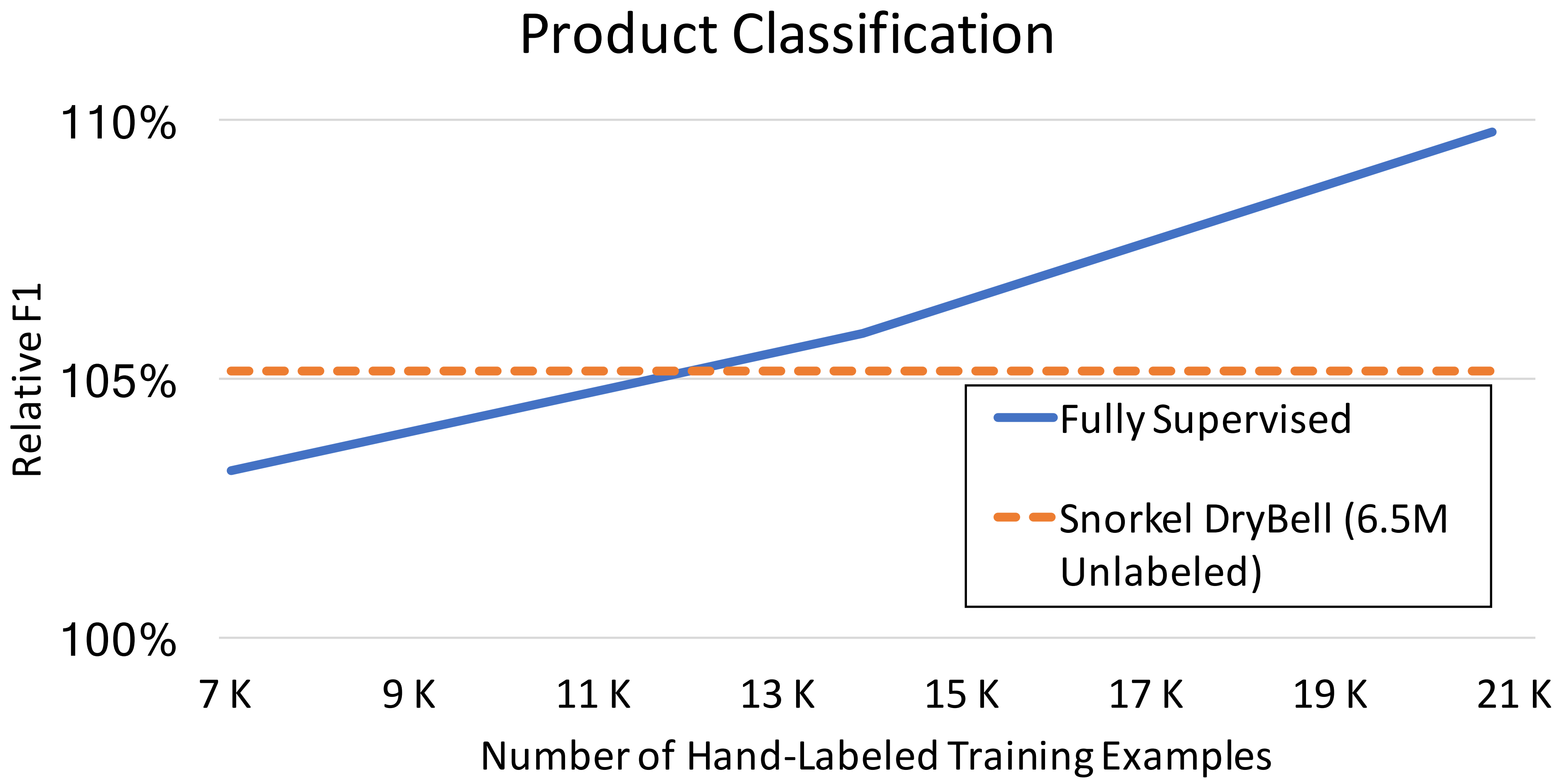
~10k labeled validation set

~10k labeled test set

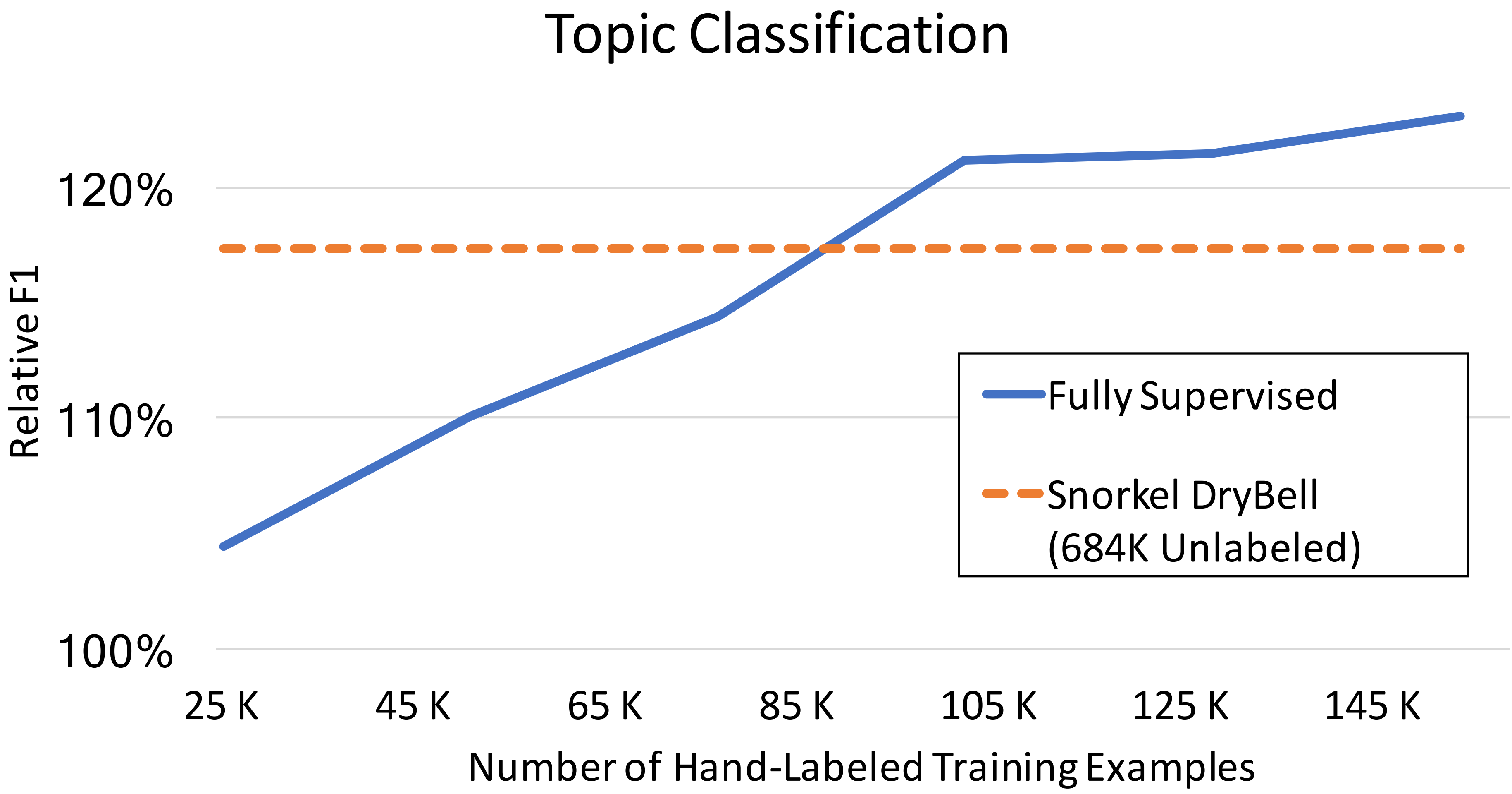
Comparison with Baselines

	Products		Topics	
	Rel. F1	Lift	Rel. F1	Lift
Train on Val. Data	100%		100%	
Generative Model	103%	+3%	94%	-6%
Snorkel DryBell	105%	**+5%**	118%	**+18%**

Break-Even Point



Break-Even Point



Summary

Summary

- Snorkel DryBell is a **new system for industrial workloads**, enabling users to transfer knowledge from organization resources to machine learning classifiers

- Our study shows that Snorkel DryBell can **save labeling tens of thousands of training examples**

- The key lesson for other organizations: **knowledge resources are abundant**, take advantage of them!

More Information

snorkel.stanford.edu

Snorkel DryBell: A Case Study in Deploying Weak Supervision at Industrial Scale. S. H. Bach, et al. SIGMOD 2019 Industrial Track. https://arxiv.org/abs/1812.00417

Thank you!

Appendix

Snorkel DryBell Scales Up to Big Data

- Using Google's distributed compute environment, we can, for example, label and fit the generative model for **5 million+ examples** in **~30 minutes**.

- Scalability of the **generative model** relies on new, **TensorFlow-based implementation**

TensorFlow

Non-Servable Resources

	Products		Topics	
	Rel. F1	Lift	Rel. F1	Lift
Servable Resources	63%		86%	
+ Non-Servable	105%	**+68%**	118%	**+36%**

Table 1: Number of unlabeled examples used during training n, number of labeled examples in the development set n_{Dev} and test set n_{Test}, fraction of positive labels in n_{Test}, and number of labeling functions used for each task, for the content classification applications.

Task	n	n_{Dev}	n_{Test}	% Pos.	# LFs
Topic Classification	684K	11K	11K	0.86	10
Product Classification	6.5M	14K	13K	1.48	8

Labeling Function Details: Topic

- 10 labeling functions
- Examples:
  - URL-based: Heuristics regarding URLs in the content
  - NER tagger-based: Heuristics based on presence of named entities
  - Topic model-based: Heuristics based on coarse-grain topic model

Labeling Function Details: Product

- 8 labeling functions

- Examples:
  - Keyword-based: rules looking for product-related keywords

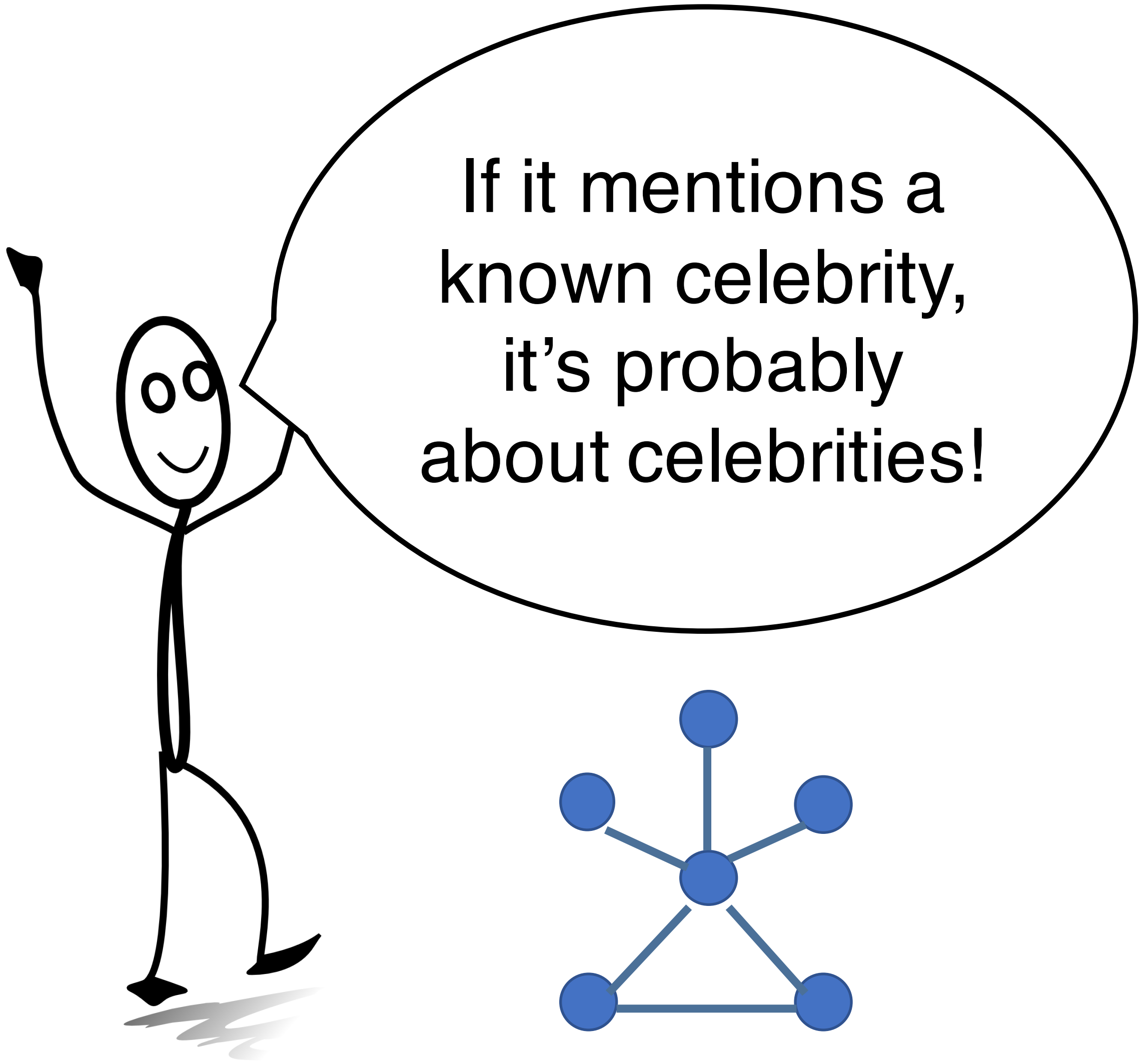
  - Knowledge Graph-based: queried for names of related products and translations in 10 languages for which the classifier is used

  - Topic model-based: Heuristics based on coarse-grain topic model

Table 4: An ablation study of `Snorkel DryBell` using equal weights for all labeling functions to label training data ("Equal Weights") compared with using the weights estimated by the generative model. All scores are normalized to the precision, recall, and F1 of the logistic regression classifier trained directly on the development set. Lift is reported relative to Equal Weights.

Relative:	P	R	**F1**	**Lift**
Topic Classification				
Equal Weights	54.1%	163.7%	109.0%	
+ Generative Model	100.6%	132.1%	117.5%	*+7.7%*
Product Classification				
Equal Weights	94.3%	110.9%	103.24%	
+ Generative Model	99.2%	110.1%	105.2%	*+1.9%*

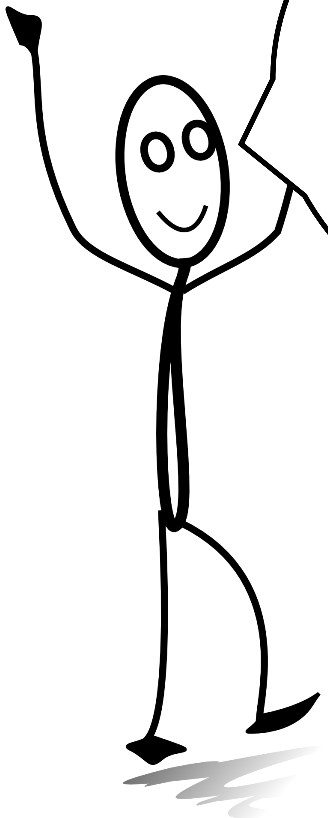
Example 2: Knowledge Graph



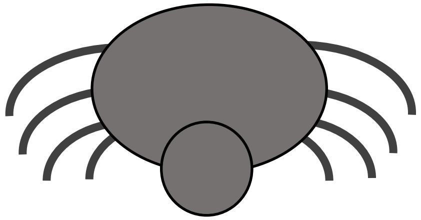
Vote

Headline	Vote
Lori Loughlin's 'Fuller House' Future Dim Due To Elite School Bribery Scandal	✓
HOW TO WATCH TESLA'S MODEL Y REVEAL TONIGHT	?
Morris Dees, a Co-Founder of the Southern Poverty Law Center, Is Ousted	?

Example 3: Web Crawler



If it points to a page that mentions lots of celebrities, it is probably about celebrities!



Vote

Lori Loughlin's 'Fuller House' Future Dim Due To Elite School Bribery Scandal ✓

HOW TO WATCH TESLA'S MODEL Y REVEAL TONIGHT ✓

Morris Dees, a Co-Founder of the Southern Poverty Law Center, Is Ousted ?