Human-Centered Machine Learning

Saleema Amershi
Machine Teaching Group, Microsoft Research

UW CSE 510 Lecture, March 1, 2016

What is Machine Teaching?

Can improve learning with better learning strategies:

- Note taking
- Self-explanation
- Practice
- Mnemonic devices
- ….

Machine Learning

Machine Teaching

Images from http://thetomatos.com

What is Machine Learning?

What is Machine Learning?

“Process by which a system improves performance from experience.” – Herbert Simon

“Study of algorithms that improve their performance P at some task T with experience E” – Tom Mitchell

“Field of study that gives computers the ability to learn without being explicitly programmed” – Arthur Samuel

Programming

6
5
3
1
8
7
2

→

→

1
2
3
5
6
7
8

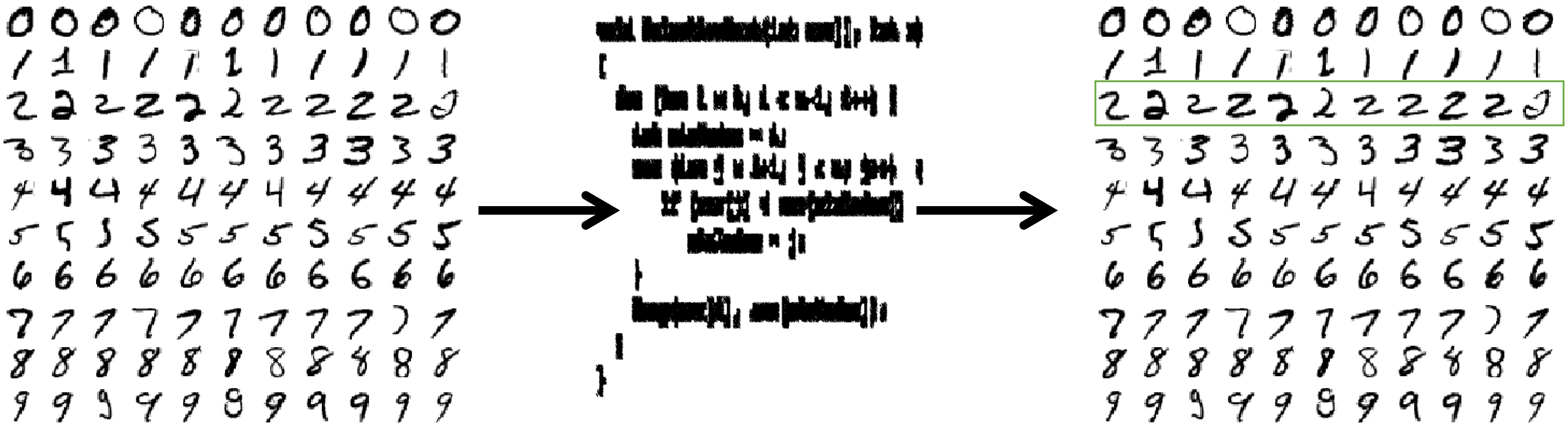
Programming

6
5
3
1
8
7
2

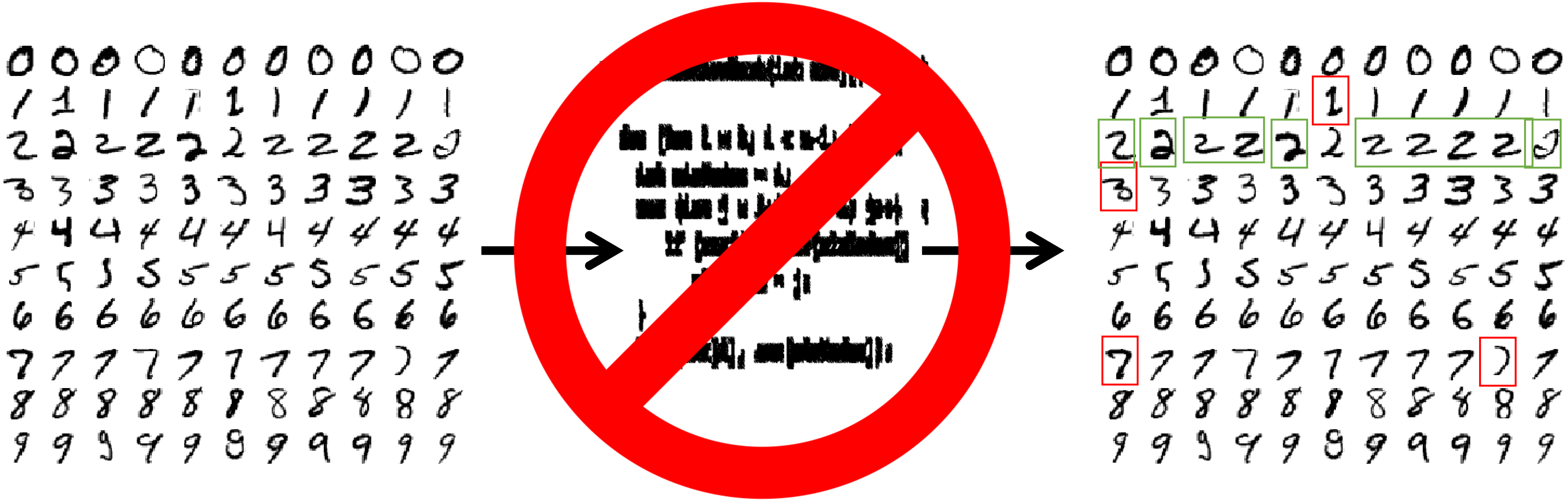
```
void SelectionSort(int arr[], int n)
{
  for (int i = 0; i < n-1; i++) {
    int minIndex = i;
    for (int j = i+1; j < n; j++)  {
       if (arr[j] < arr[minIndex])
          minIndex = j;
    }
    Swap(arr[i], arr[minIndex]);
  }
}
```

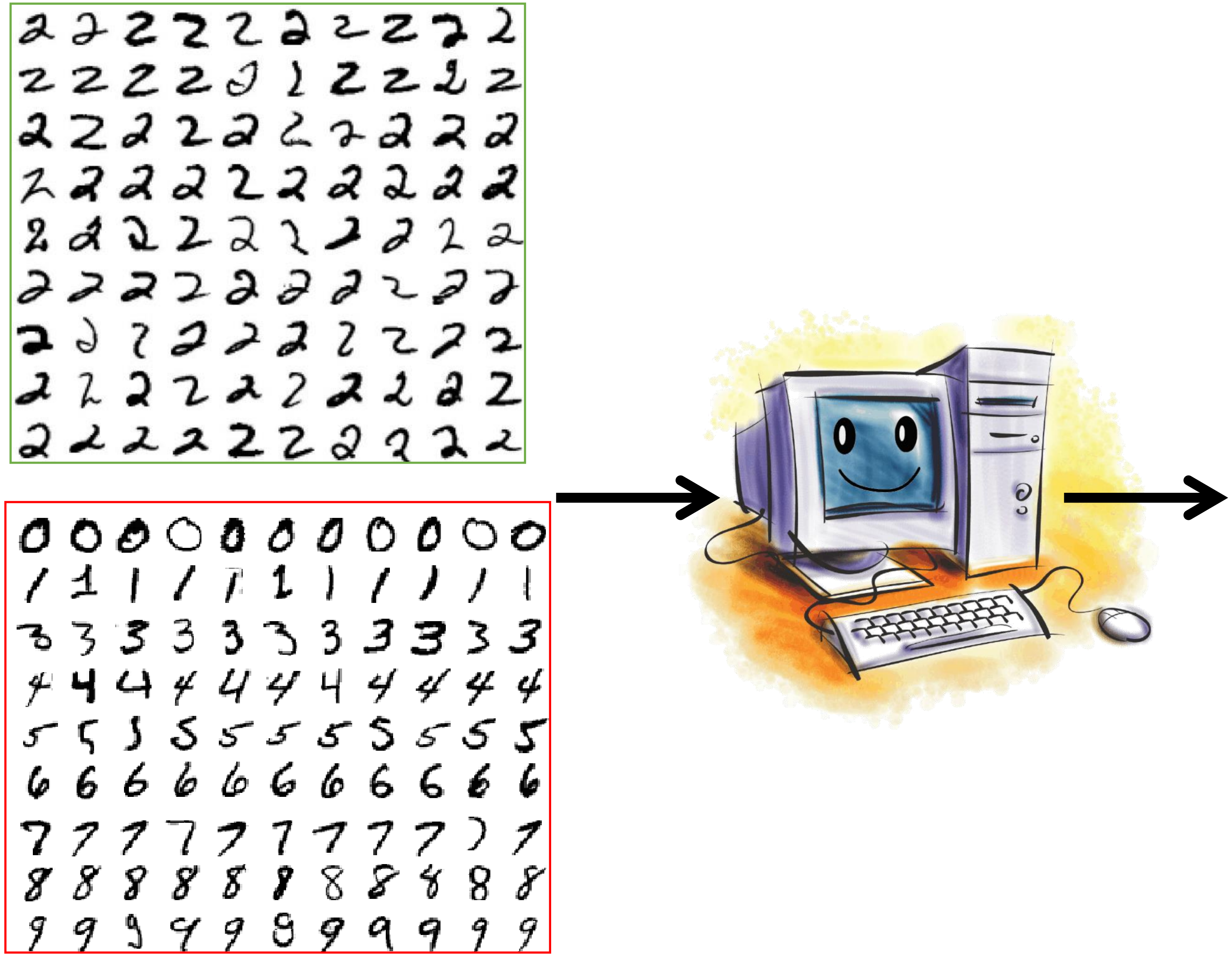
1
2
3
5
6
7
8

Programming

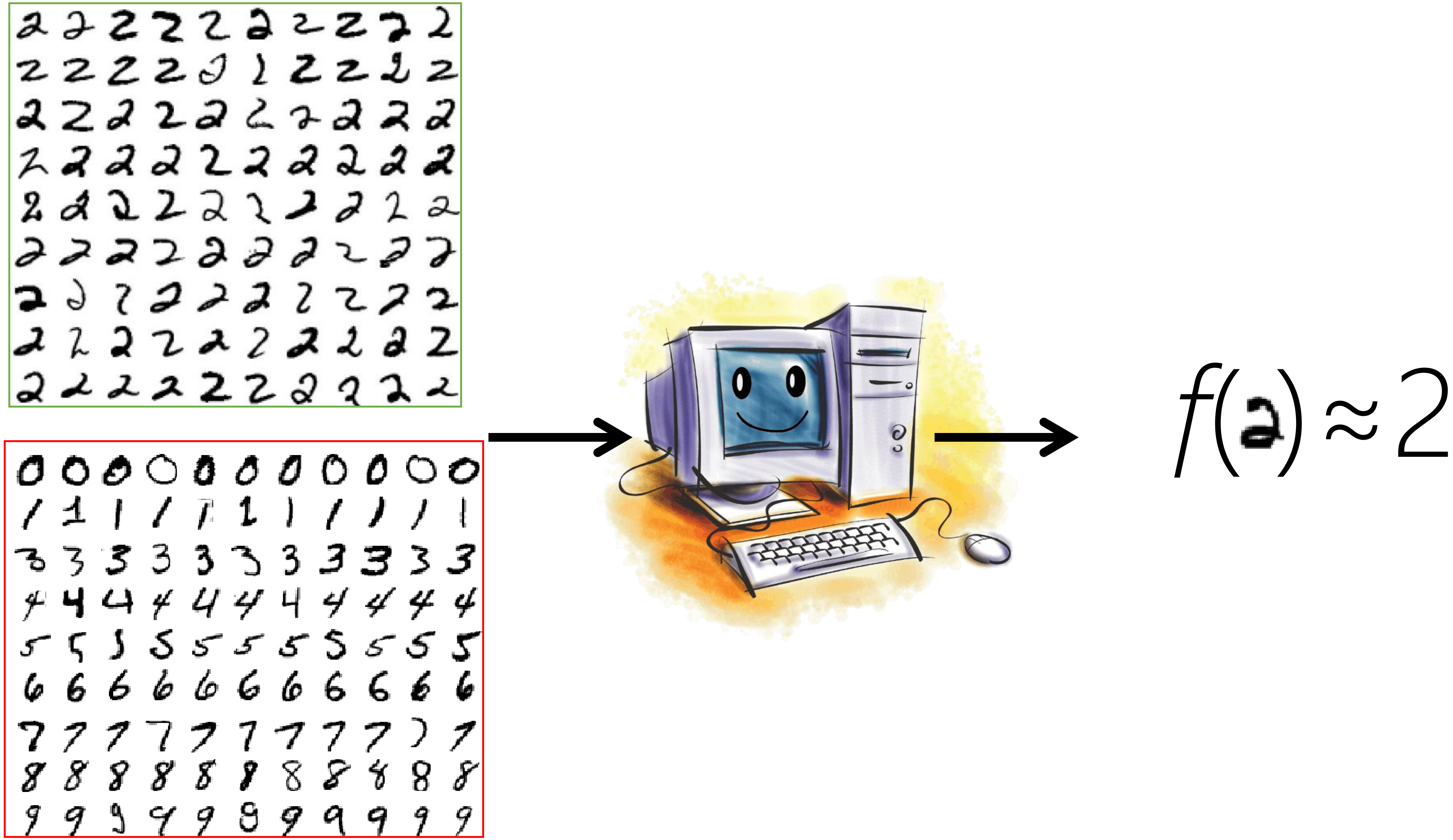


Programming



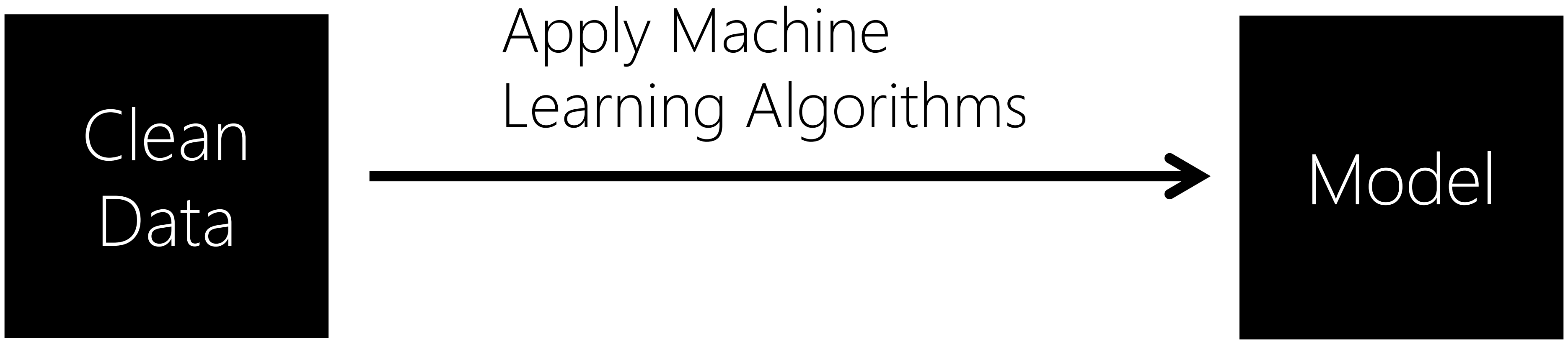


$$f(x) \approx y$$

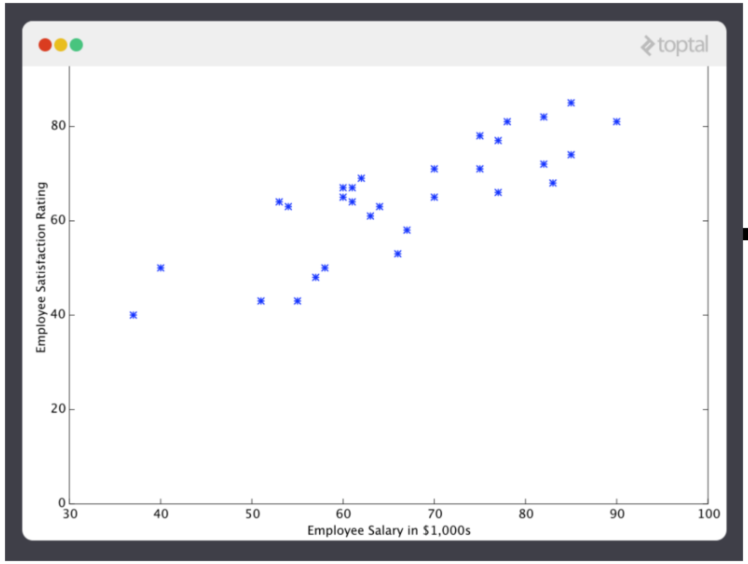

$f(2) \approx 2$

How do Machines Learn?

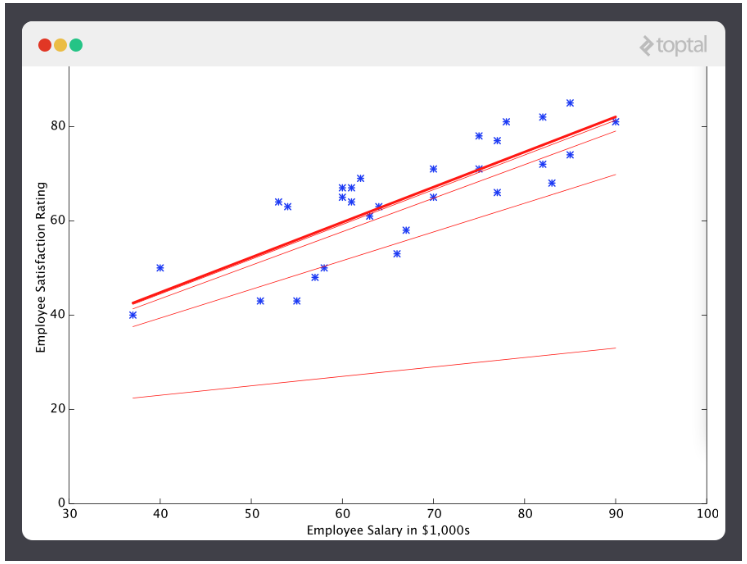
How do Machines Learn?

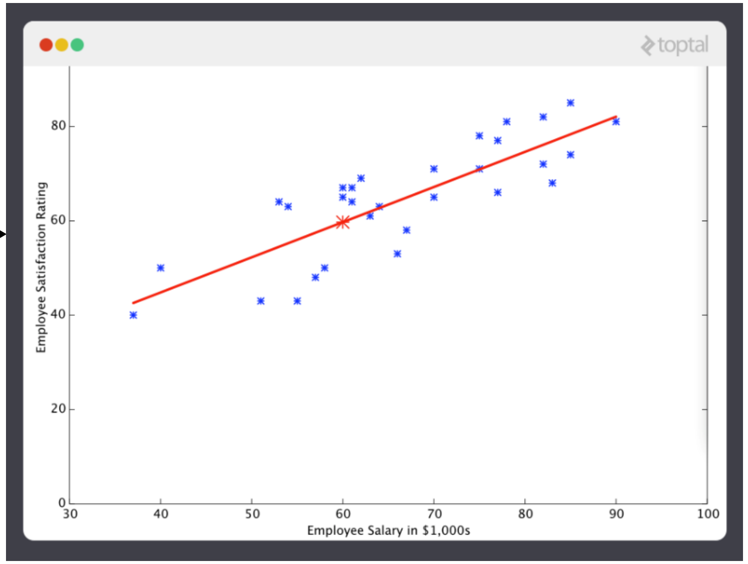


How do Machines Learn?



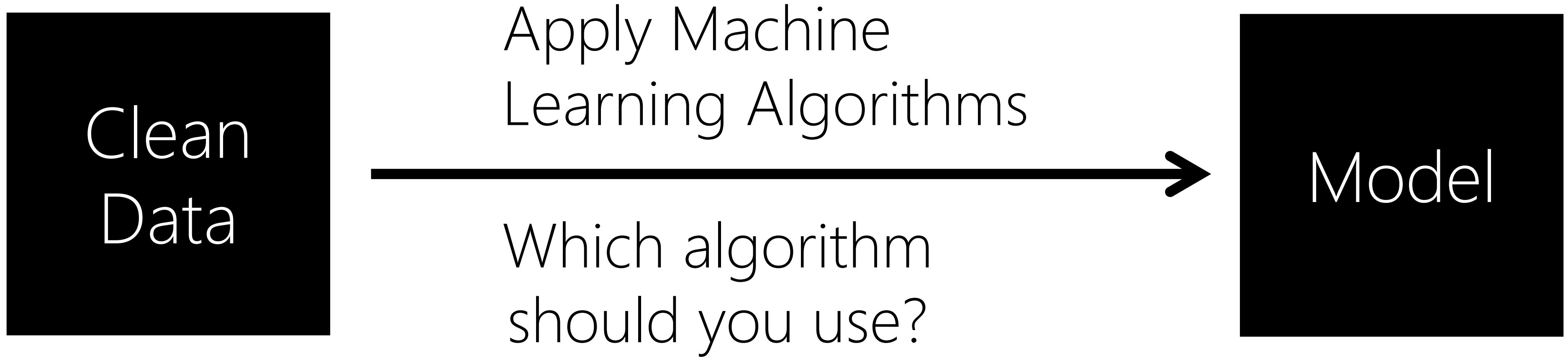
Apply Machine
Learning Algorithms





Images from: https://www.toptal.com/machine-learning/machine-learning-theory-an-introductory-primer

How do Machines Learn?



Where do you get this data?
How should it be represented?

How do you know if its working?

Investigating Statistical Machine Learning as a Tool for Software Developers

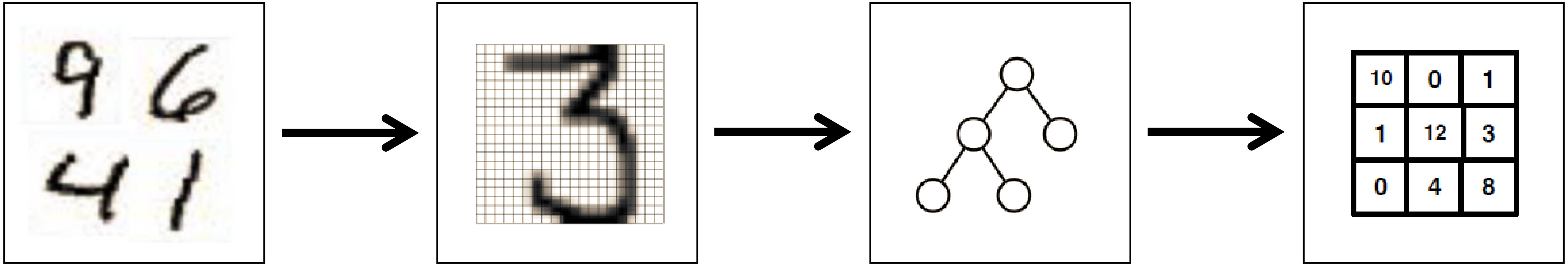
Patel, K., Fogarty, J., Landay, J., and Harrison, B. CHI 2008.

Methodology

Semi-structured interviews with 11 researchers.

5 hour think-aloud study with 10 participants. Digit recognition task.

Applied Machine Learning is a Process



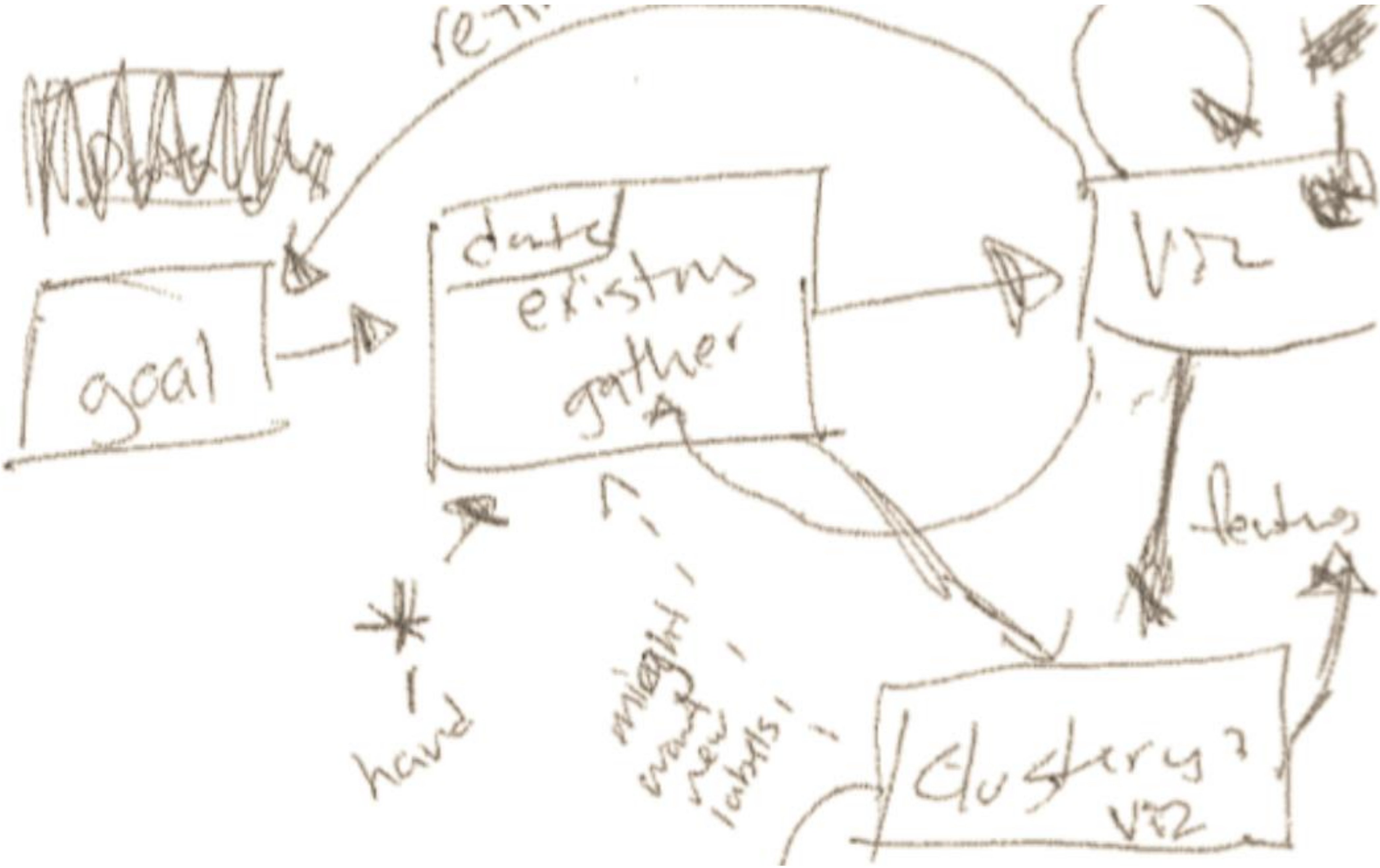
Collect Data

Create Features

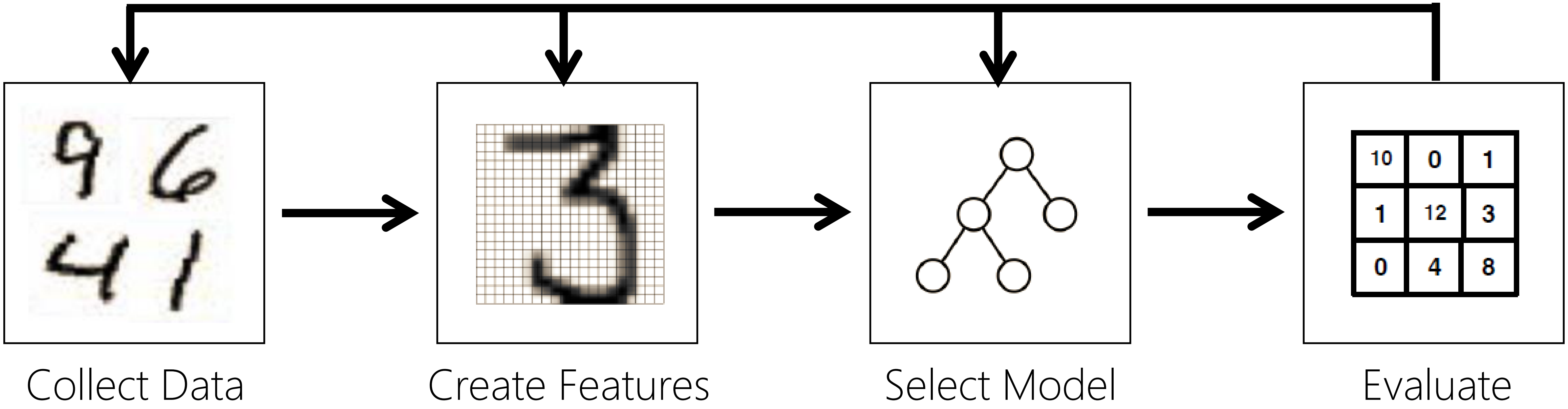
Select Model

Evaluate

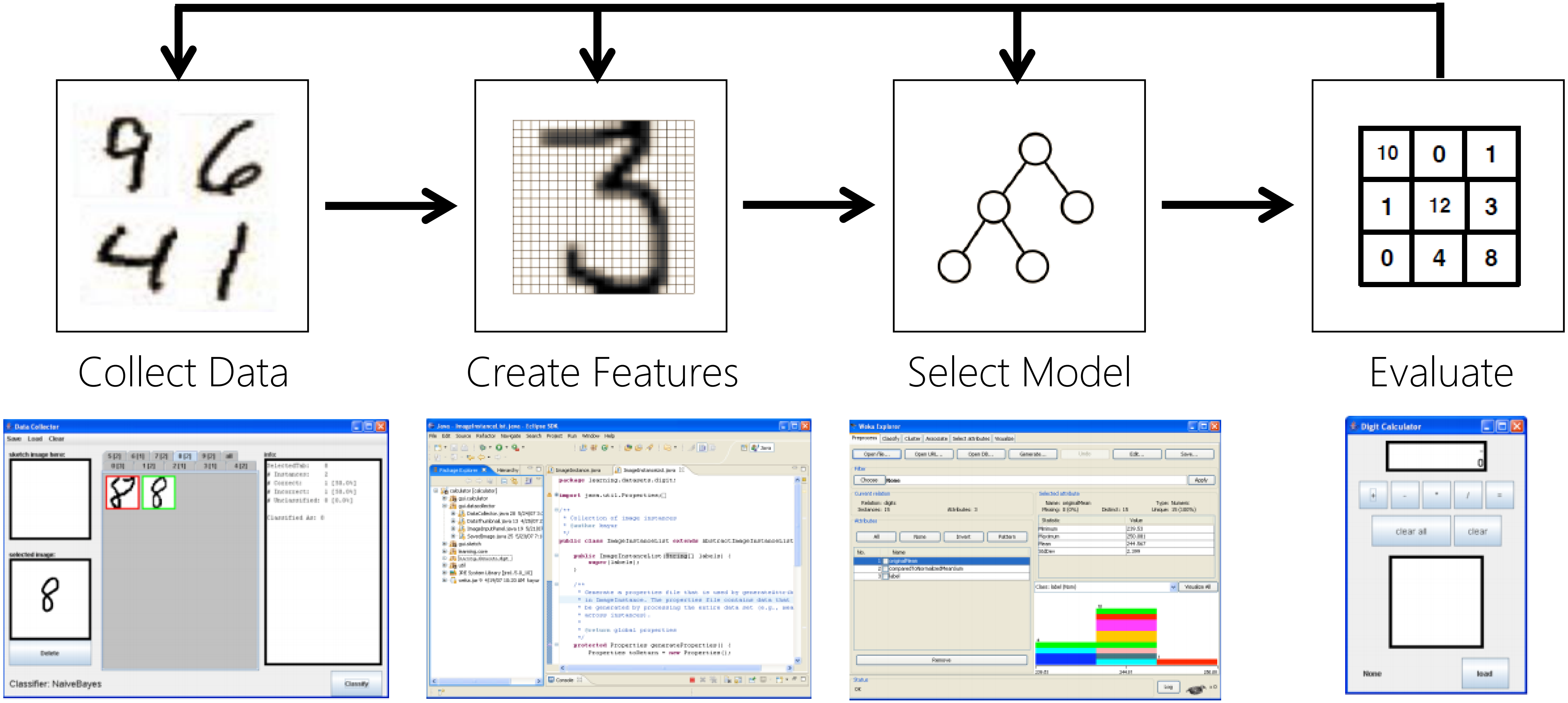
Slide content from Kayur Patel


goal
data
existing
gather
V?2
hand
might want new labels
clustery?
V?2
features

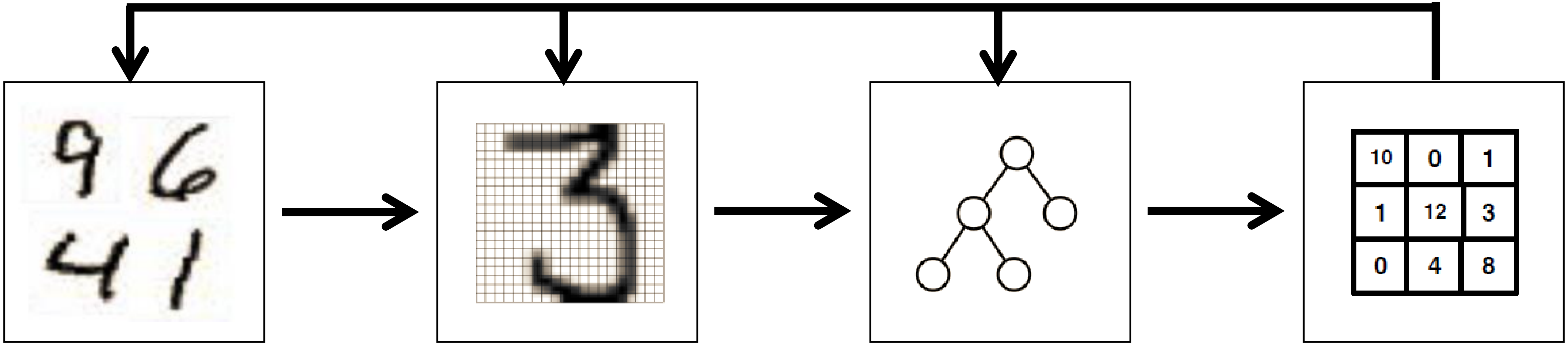
Applied Machine Learning is a Process



Applied Machine Learning is a Process

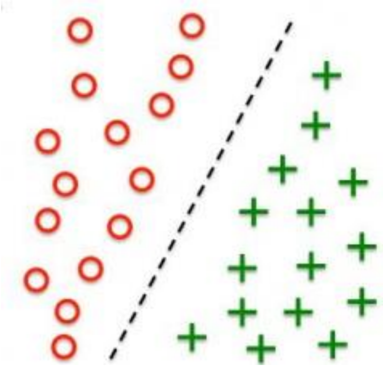


What About Music Recommendation?



Collect Data | Create Features | Select Model | Evaluate

Genre: Rock
Tempo: Fast
Drums: Yes
Time of day: Afternoon
Recently heard: No
....



Problems with current tools

Don’t support machine learning as an iterative and exploratory process.

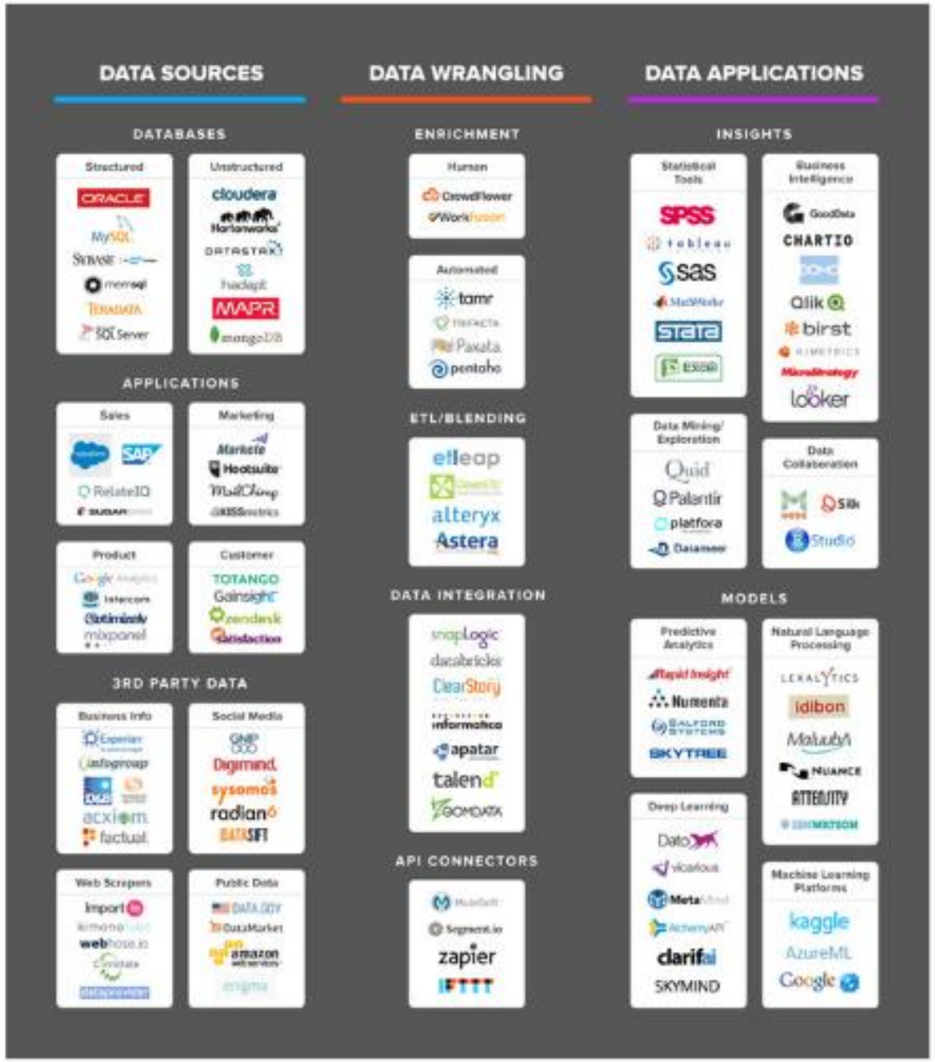
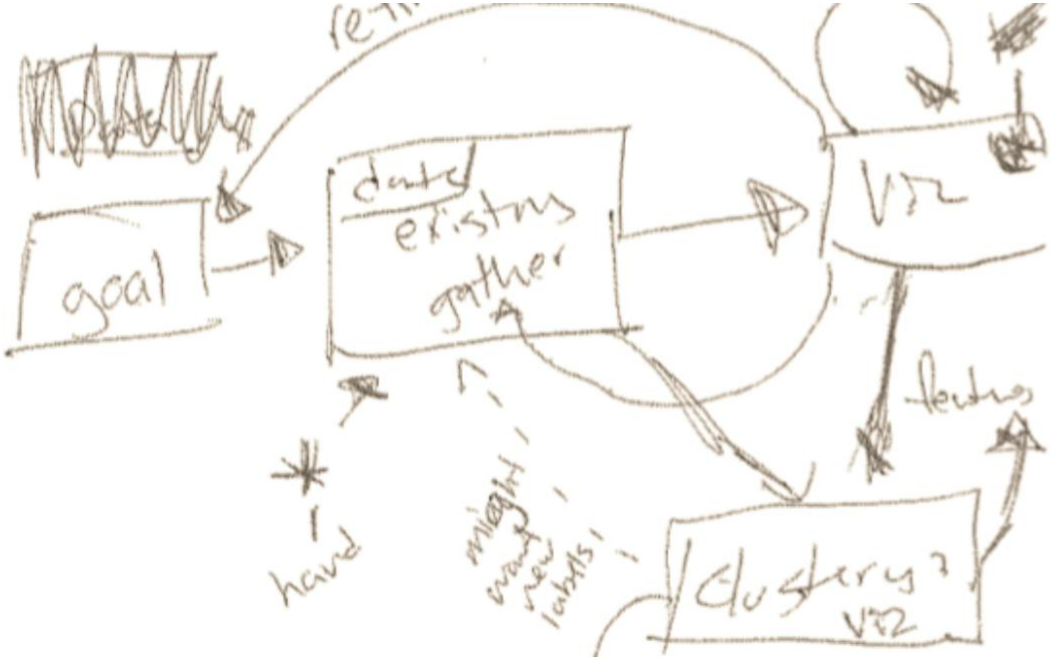


Image from: http://www.crowdflower.com/blog/the-data-science-ecosystem

Problems with current tools

Don’t support machine learning as an iterative and exploratory process.

Don’t support relating data to behaviors of the algorithm.



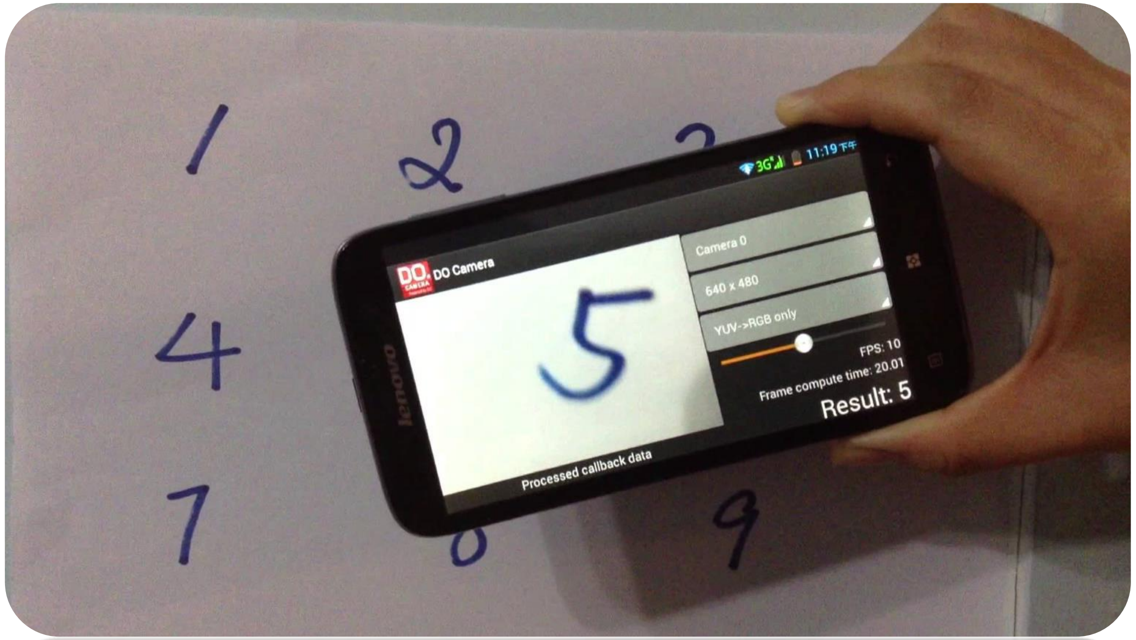
LogitBoostWith8To18EvenWindow-Iter=10.model
LogitBoostWith8To18EvenWindow-Iter=20.model
SVMWith8To18EvenWindow-Iter=10.model
....

Problems with current tools

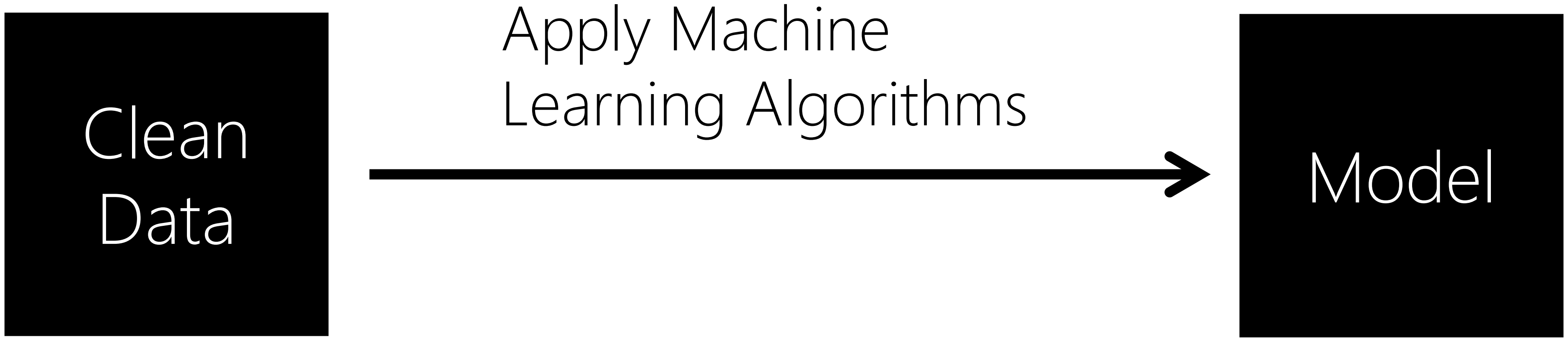
Don't support machine learning as an iterative and exploratory process.

Don't support relating data to behaviors of the algorithm.

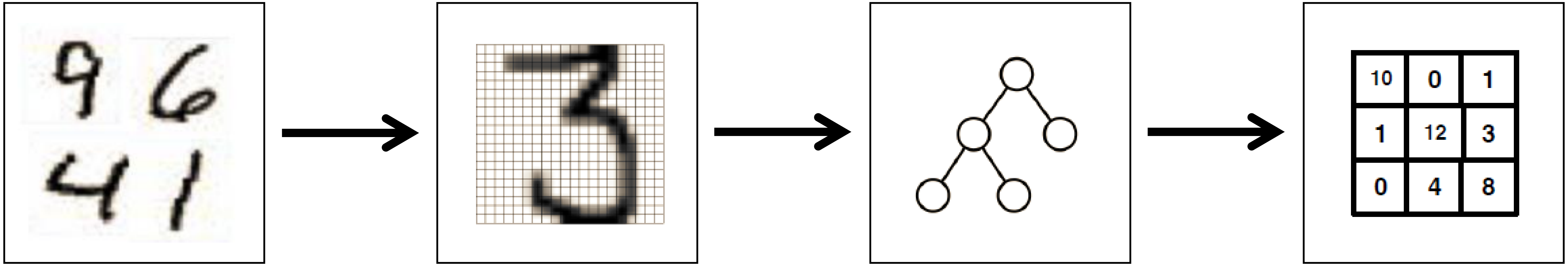
Don't support evaluation in context of use.



Model Performance is Important

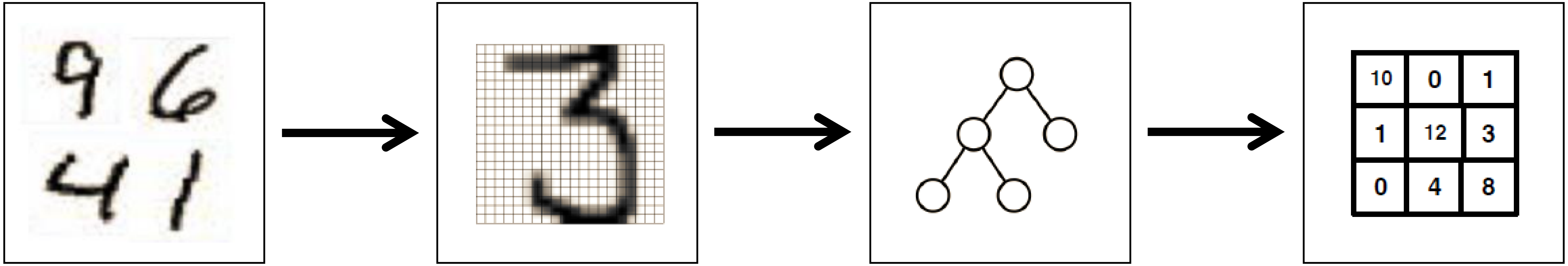


What are other considerations?

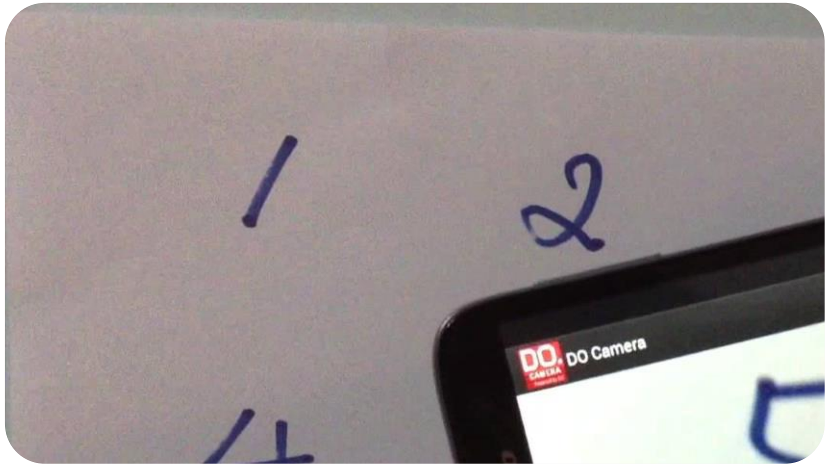


Collect Data → Create Features → Select Model → Evaluate

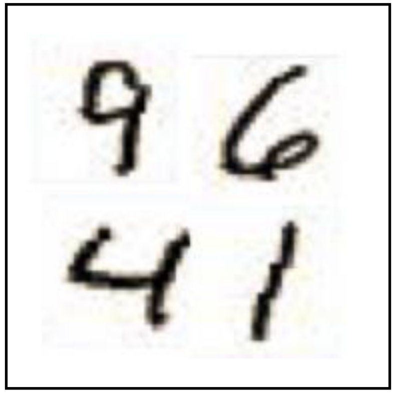
Computational Efficiency?



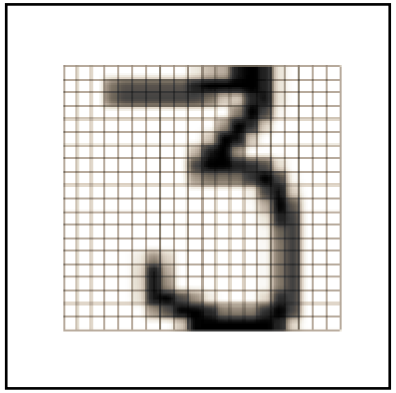
Collect Data   Create Features   Select Model   Evaluate



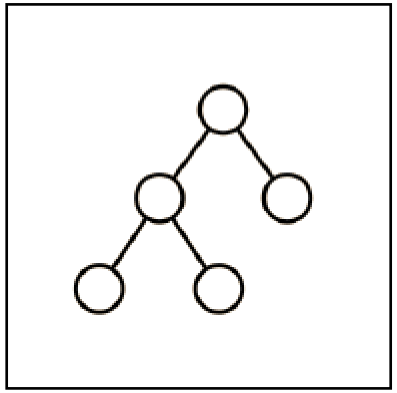
Data processing efficiency?



Collect Data



Create Features



Select Model

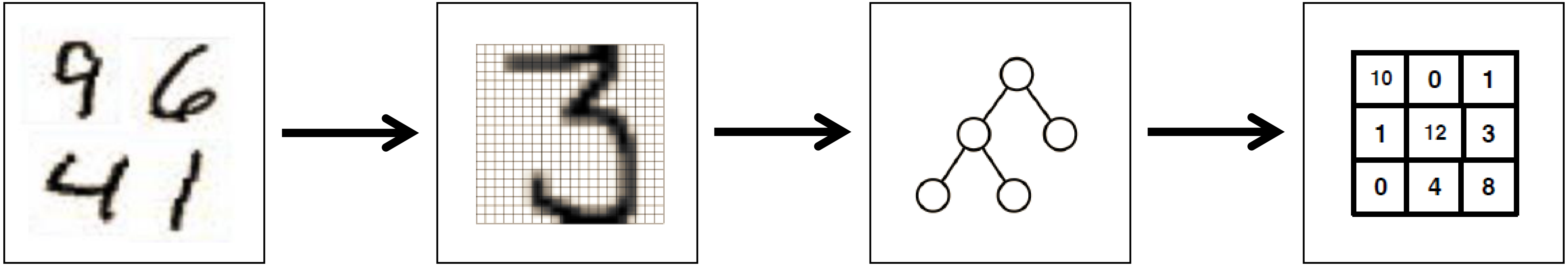
10	0	1
1	12	3
0	4	8

Evaluate

"Data scientists, according to interviews and expert estimates, spend ***50 percent to 80 percent of their time*** *mired in this more mundane labor of collecting and preparing unruly digital data."*

\- *New York Times, 2014*

Understandability?



Collect Data | Create Features | Select Model | Evaluate

"TAP9 initially used a decision tree algorithm because it allowed TAP9 to easily see what features were being used...Later in the study...they transitioned to using more complex models in search of increased performance."

Considerations for Machine Learning

Model performance

Computational efficiency

Iteration efficiency

Ease of experimentation

Understandability

....

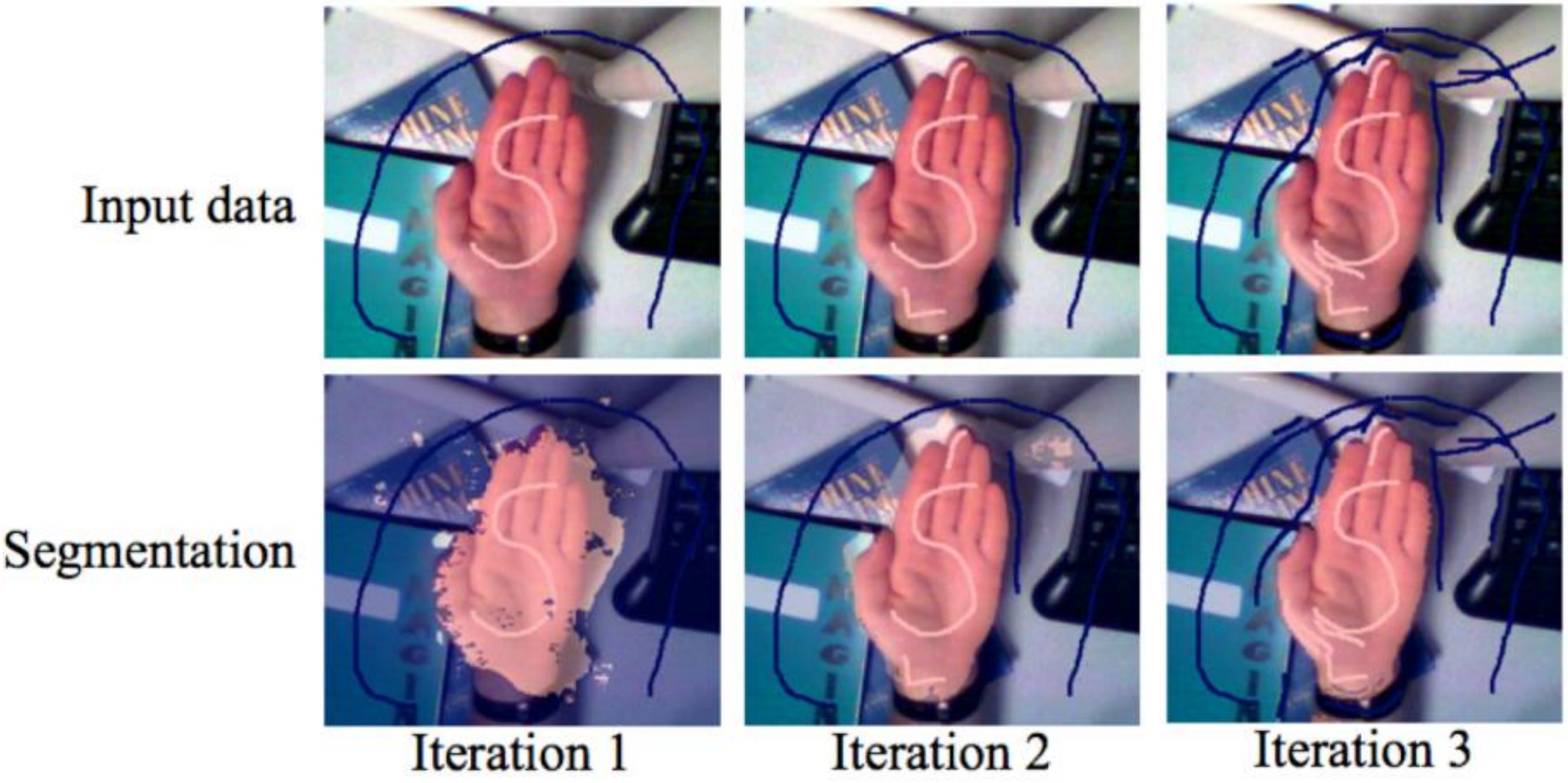
New opportunities for HCI research!

Need to make tradeoffs!

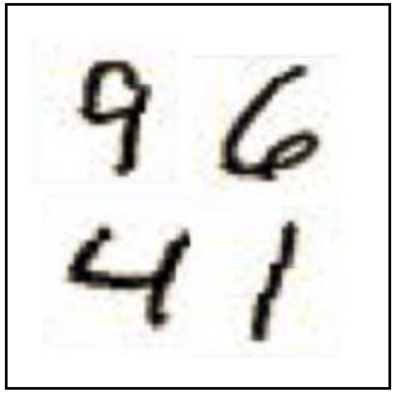
Interactive Machine Learning

Fails, J.A. and Olsen, D.R. IUI 2003.

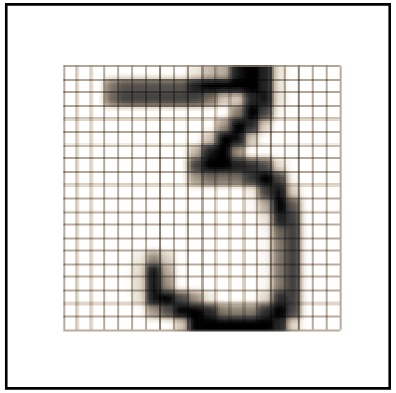
Crayons: IML for Pixel Classifiers



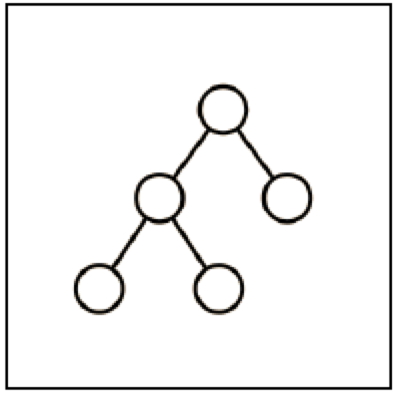
What tradeoffs did Crayons make?



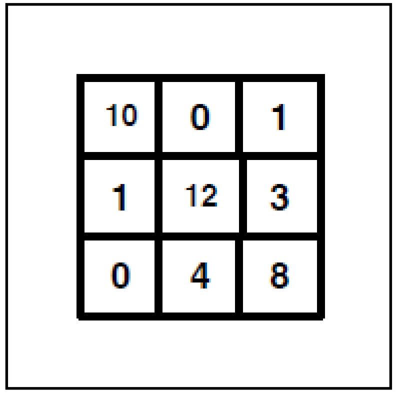
Collect Data



Create Features

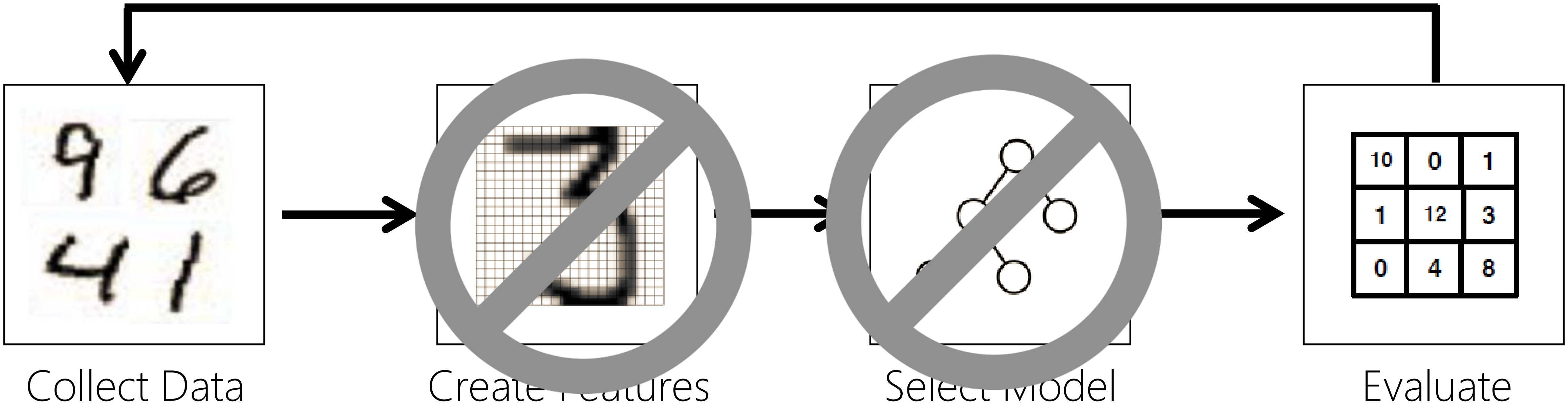


Select Model

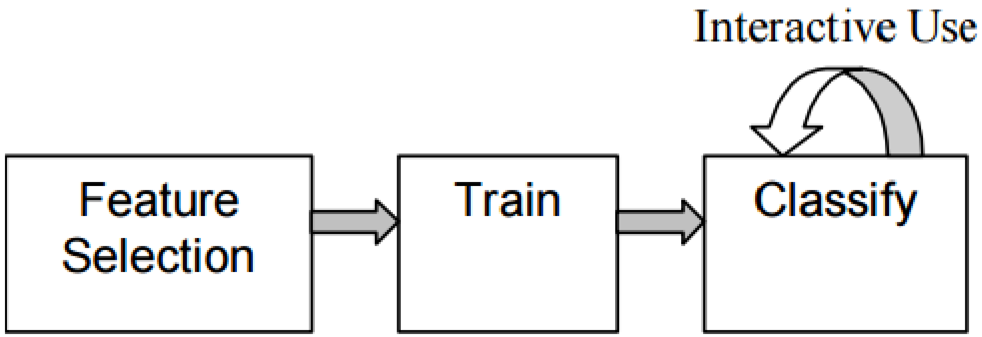


Evaluate

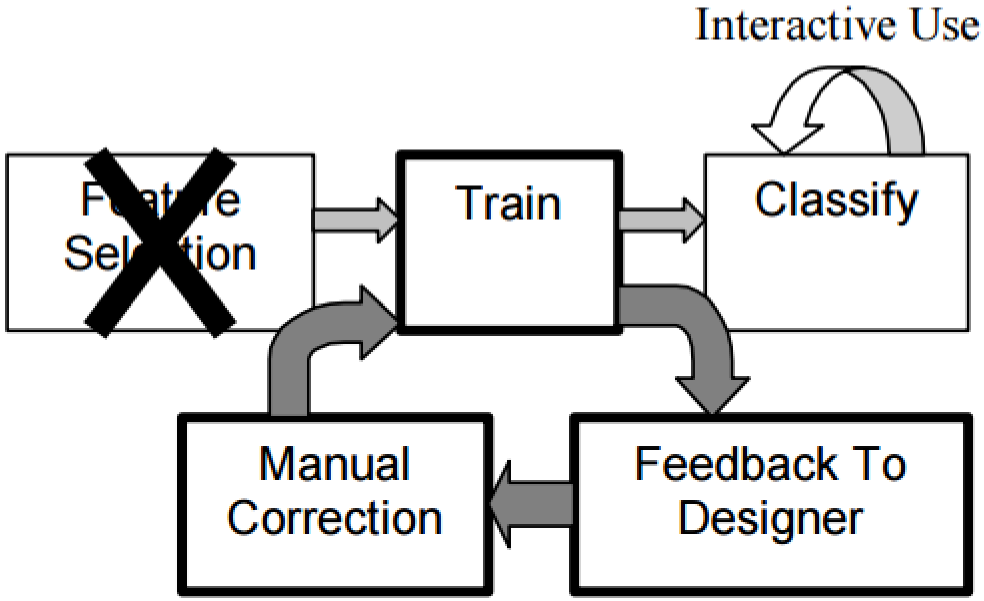
What tradeoffs did Crayons make?



“Classical” ML



Interactive ML



What tradeoffs did Crayons make?

Rapid iteration

- Fast training
- Integrated environment

Simplicity

Model performance

Flexibility

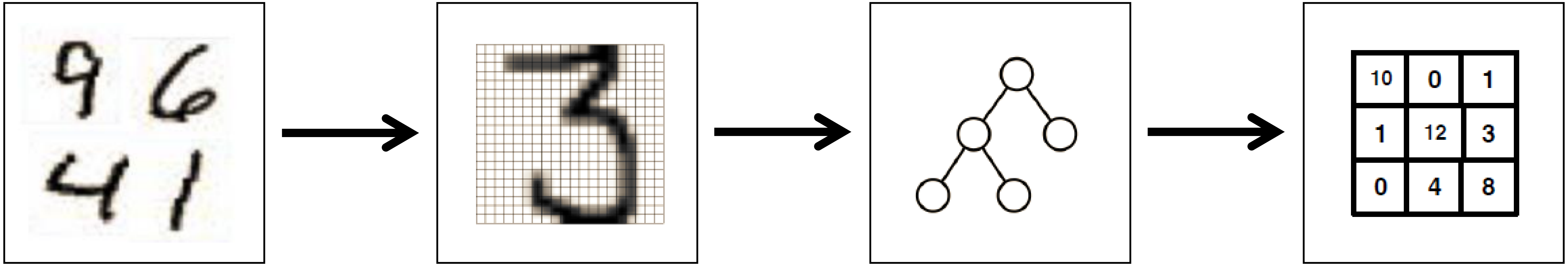
- Automatic featuring
- No model selection

When are these tradeoffs appropriate?

Rapid iteration Simplicity	Model performance Flexibility
Novices Large set of available features Data can be efficiently viewed and labeled	Experts Custom features needed Data types that can't be viewed at a glance Labels obtained from external sources

Flock: Hybrid Crowd-Machine Learning Classifiers

Cheng, J. and Bernstein, S. CSCW 2015.



Collect Data · Create Features · Select Model · Evaluate

"At the end of the day, some machine learning projects succeed and some fail. What makes the differences? ***Easily the most important factor is the features used.****"*

[Domingos, CACM 2012]

How do people come up with features?

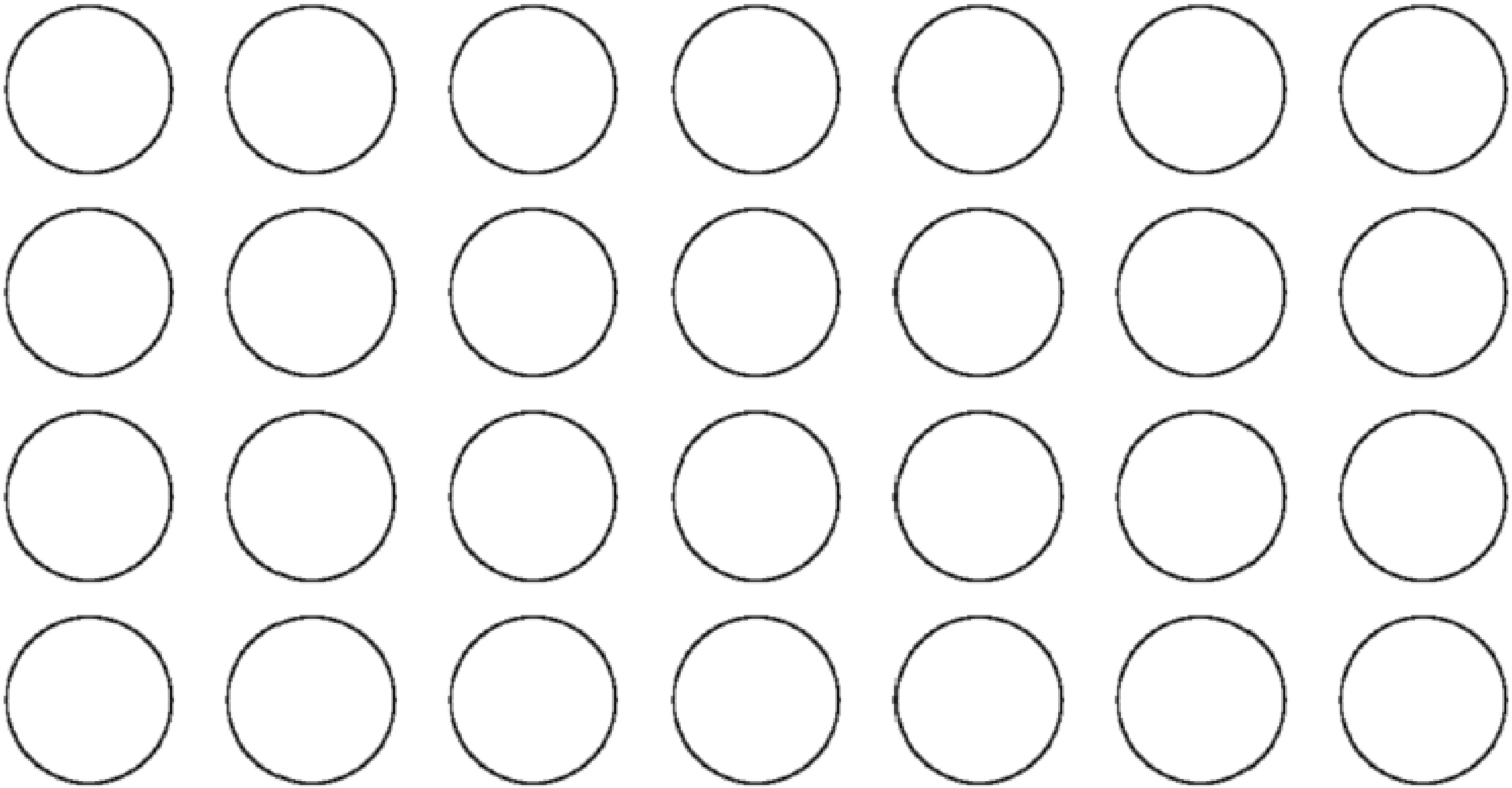
Look for features used in related domains.

Use intuition or domain knowledge.

Apply automated techniques.

Feature ideation – think of and experiment with custom features.

A Brainstorming Exercise



How do people come up with features?

Look for features used in related domains.

Use intuition or domain knowledge.

Apply automated techniques.

Feature ideation – think of and experiment with custom features.

> *"The novelty of generated ideas increases as participants ideate, reaching a peak after their 18th instance."*
>
> [Krynicki, F. R., 2014]

Workflow

User specifies a concept and uploads some unlabeled data.

Crowd views data and suggests features.

What makes a cat a cat?

What makes a cat a cat?

Workflow

User specifies a concept and uploads some unlabeled data.

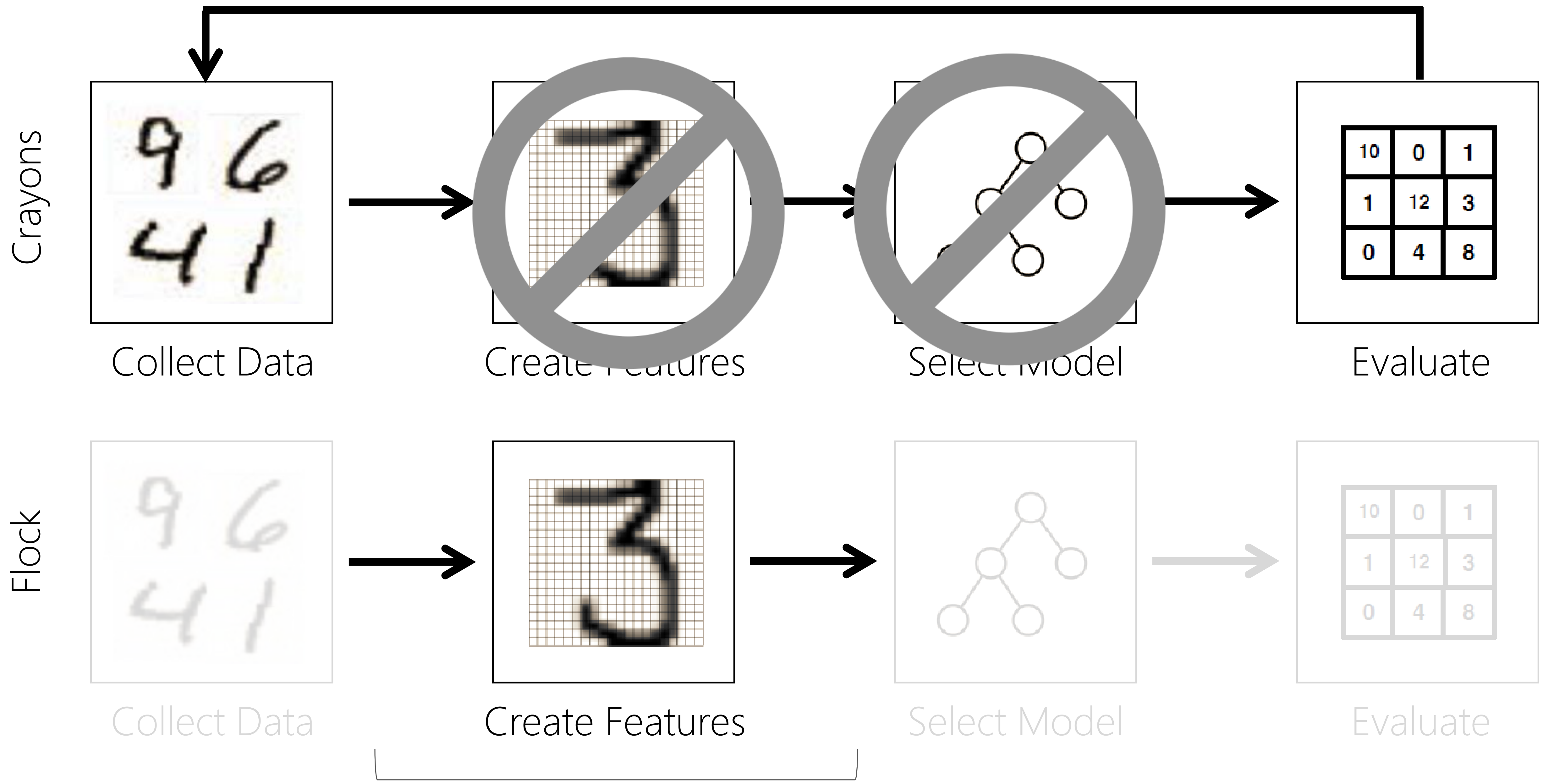
Crowd compares and contrasts positive and negative examples and suggests “why” they are different. Reasons become features.

Reasons are clustered.

User vets, edits, and adds to features.

Crowd implements feature by labeling data.

Features used to build classifiers.



Crayons
9 6
4 1
Collect Data
Create Features
Select Model
10 0 1
1 12 3
0 4 8
Evaluate
Flock
9 6
4 1
Collect Data
Create Features
Select Model
10 0 1
1 12 3
0 4 8
Evaluate

Collect Data

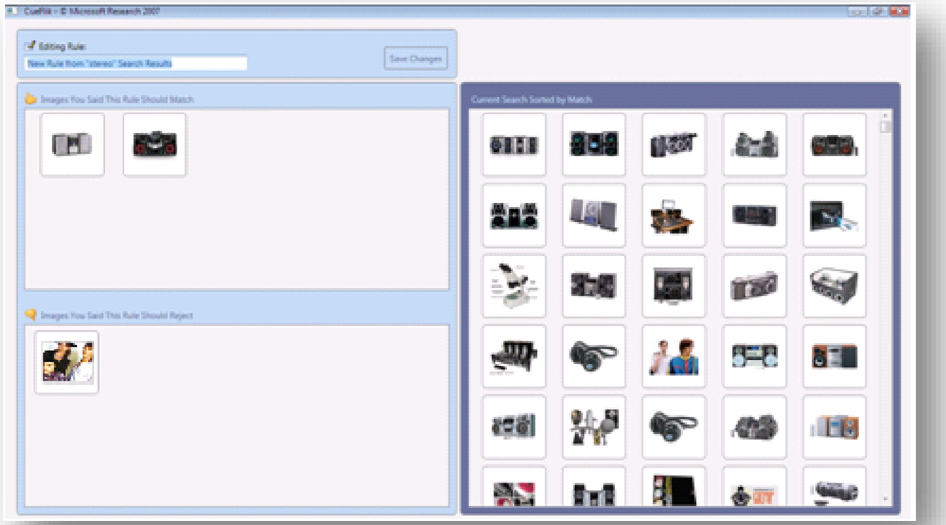
Create Features

Select Model

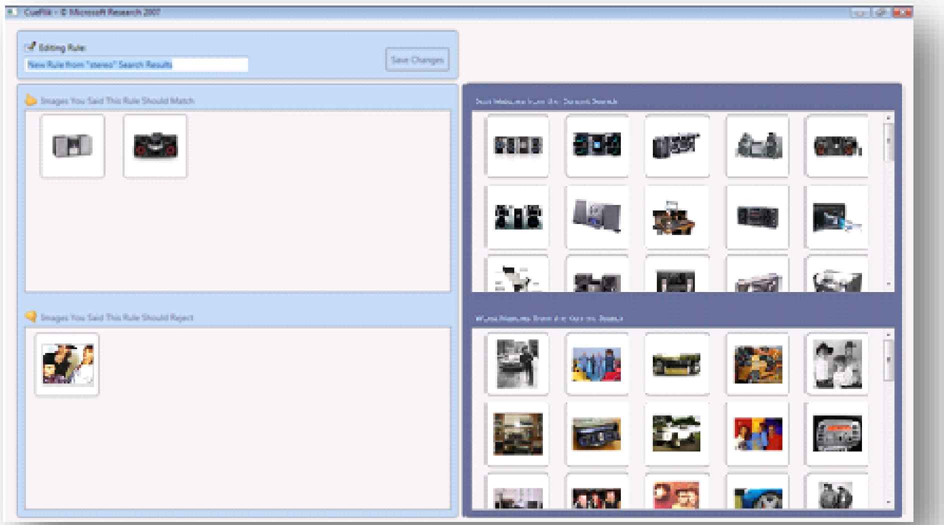
Evaluate

Positives

Negatives

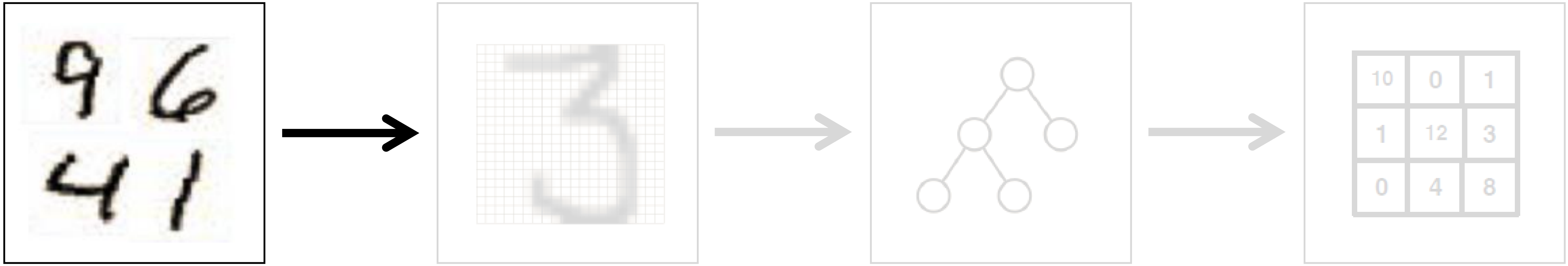


Standard Ranked List



Split Technique (Best/Worst Matches)

[Fogarty et al., CHI 2007]



Collect Data

Create Features

Select Model

Evaluate

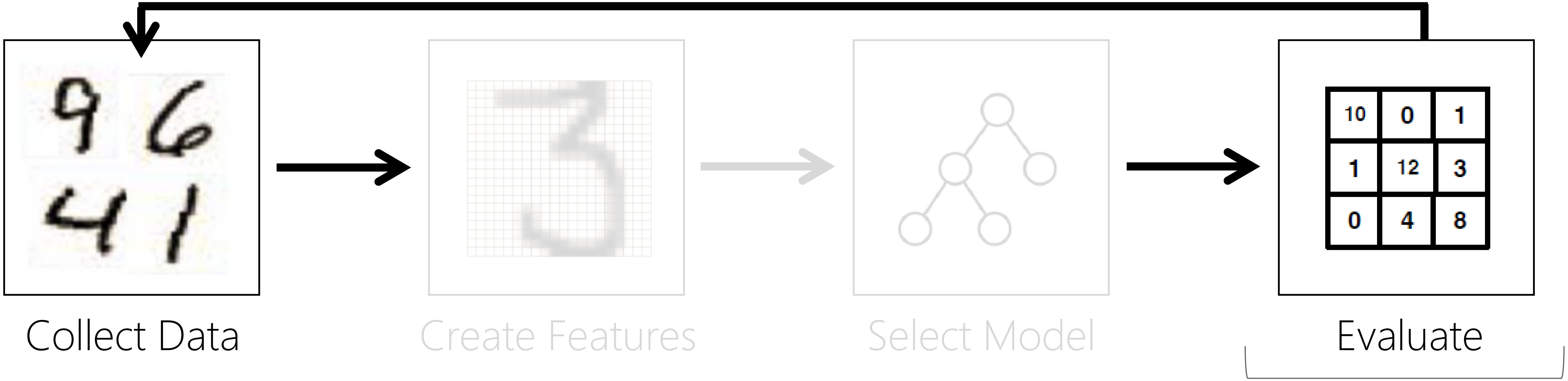
Traditional Labeling

Structured Labeling

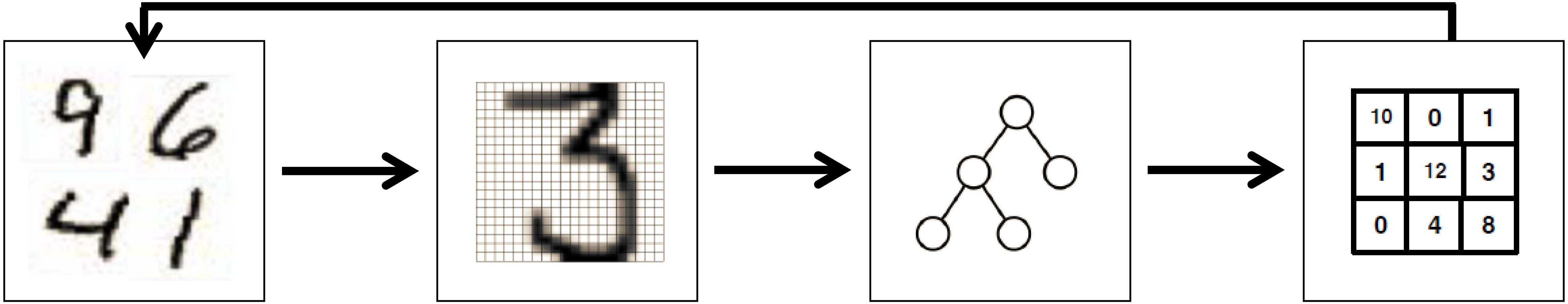
Grouping and tagging surfaces decision making.

Moving, merging and splitting groups helps with revising decisions.

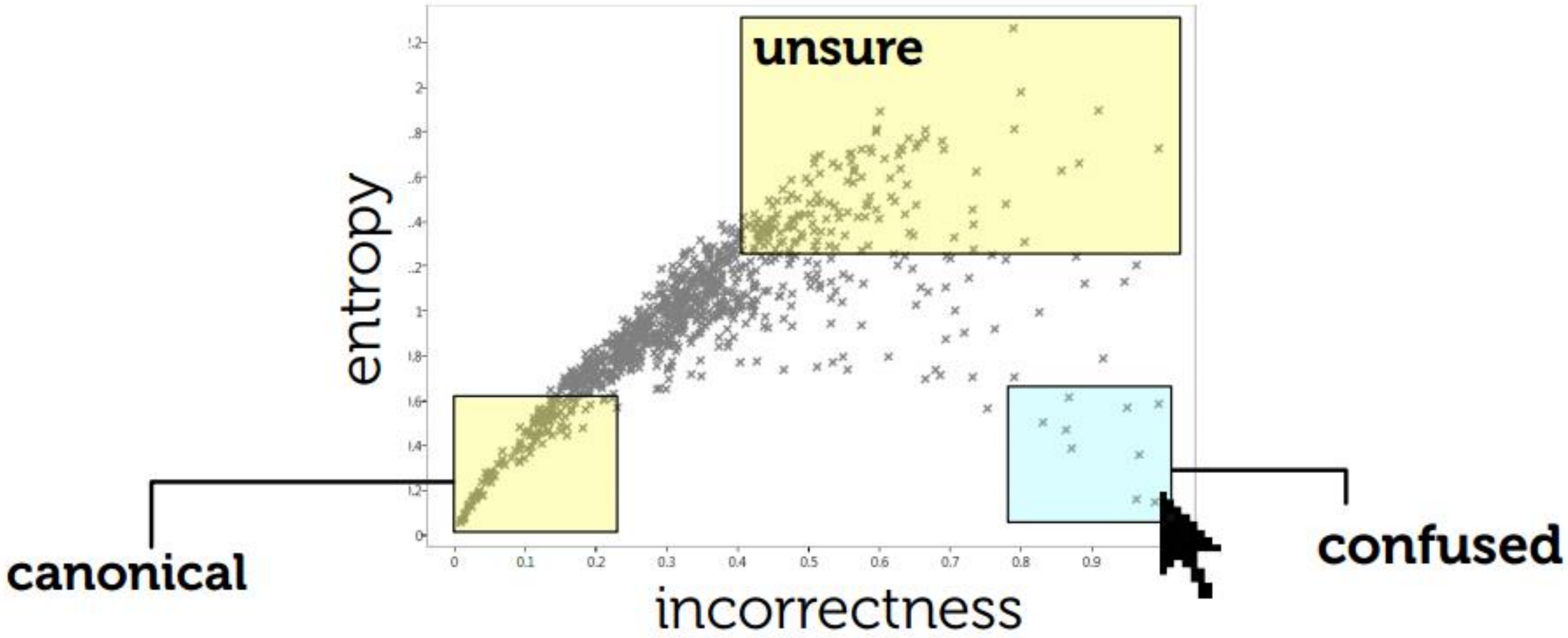
[Kulesza et al., CHI 2014]



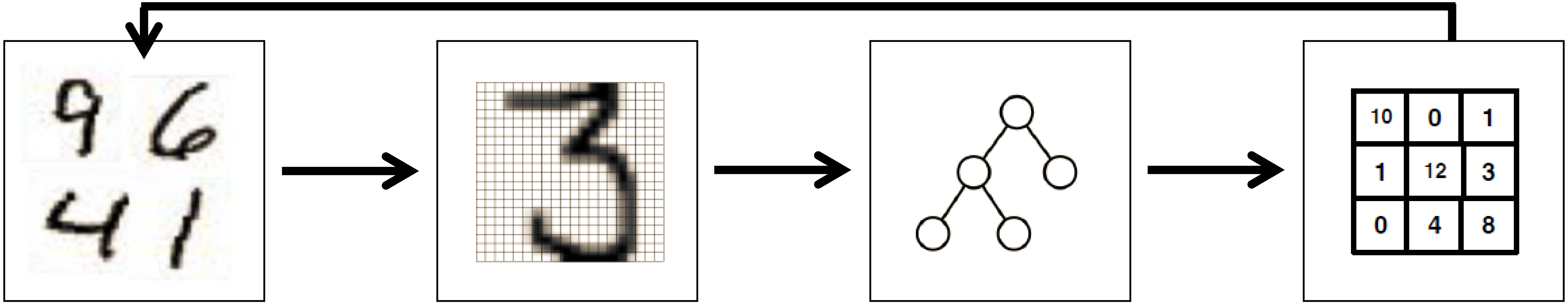
[Amershi et al., CHI 2015]



Collect Data | Create Features | Select Model | Evaluate



[Patel et al., IJCAI 2011]



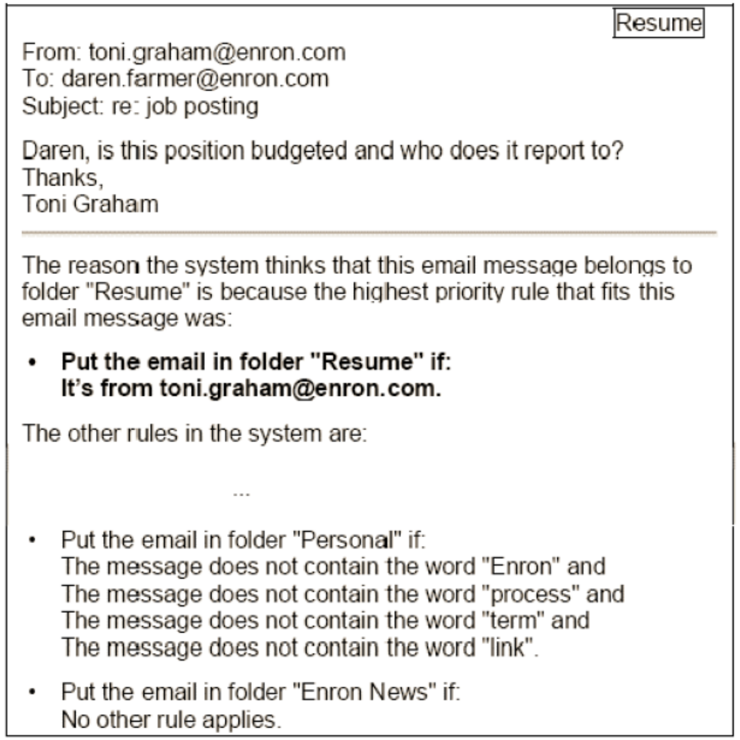
Collect Data

Create Features

Select Model

Evaluate

Rule-based explanation



Resume

From: toni.graham@enron.com
To: daren.farmer@enron.com
Subject: re: job posting

Daren, is this position budgeted and who does it report to?
Thanks,
Toni Graham

The reason the system thinks that this email message belongs to folder "Resume" is because the highest priority rule that fits this email message was:

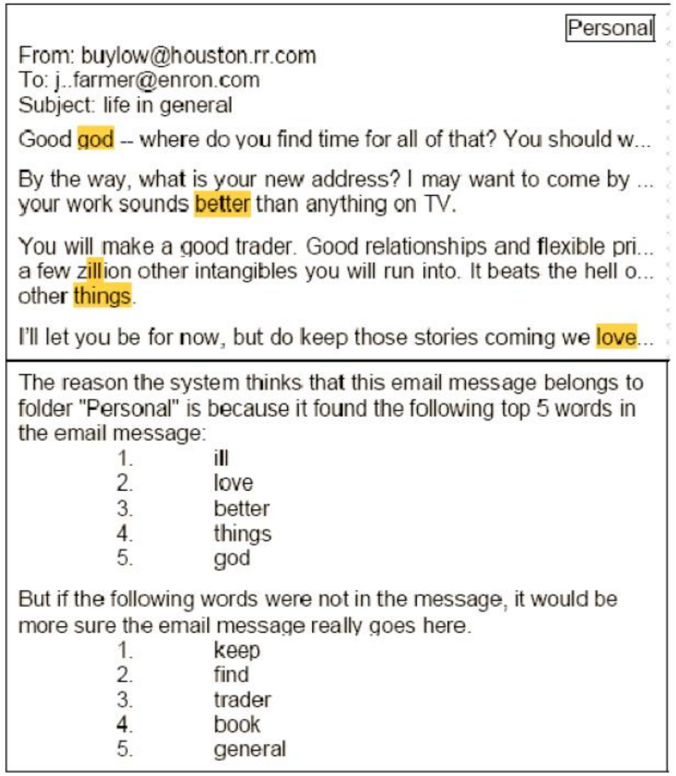
- **Put the email in folder "Resume" if: It's from toni.graham@enron.com.**

The other rules in the system are:

...

- Put the email in folder "Personal" if:
  The message does not contain the word "Enron" and
  The message does not contain the word "process" and
  The message does not contain the word "term" and
  The message does not contain the word "link".
- Put the email in folder "Enron News" if:
  No other rule applies.

Keyword-based explanation



Personal

From: buylow@houston.rr.com
To: j..farmer@enron.com
Subject: life in general

Good god -- where do you find time for all of that? You should w...

By the way, what is your new address? I may want to come by ...
your work sounds better than anything on TV.

You will make a good trader. Good relationships and flexible pri...
a few zillion other intangibles you will run into. It beats the hell o...
other things.

I'll let you be for now, but do keep those stories coming we love...

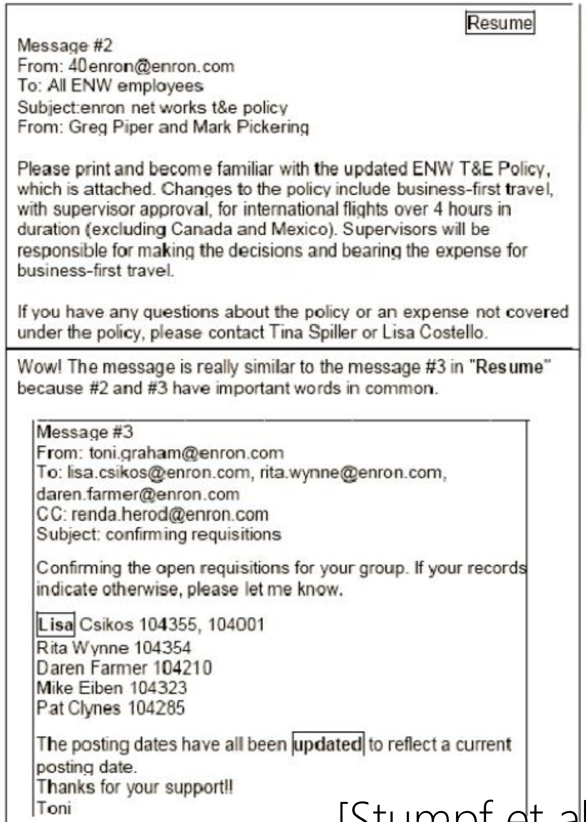
The reason the system thinks that this email message belongs to folder "Personal" is because it found the following top 5 words in the email message:

1. ill
2. love
3. better
4. things
5. god

But if the following words were not in the message, it would be more sure the email message really goes here.

1. keep
2. find
3. trader
4. book
5. general

Similarity-based explanation



Resume

Message #2
From: 40enron@enron.com
To: All ENW employees
Subject:enron net works t&e policy
From: Greg Piper and Mark Pickering

Please print and become familiar with the updated ENW T&E Policy, which is attached. Changes to the policy include business-first travel, with supervisor approval, for international flights over 4 hours in duration (excluding Canada and Mexico). Supervisors will be responsible for making the decisions and bearing the expense for business-first travel.

If you have any questions about the policy or an expense not covered under the policy, please contact Tina Spiller or Lisa Costello.

Wow! The message is really similar to the message #3 in "Resume" because #2 and #3 have important words in common.

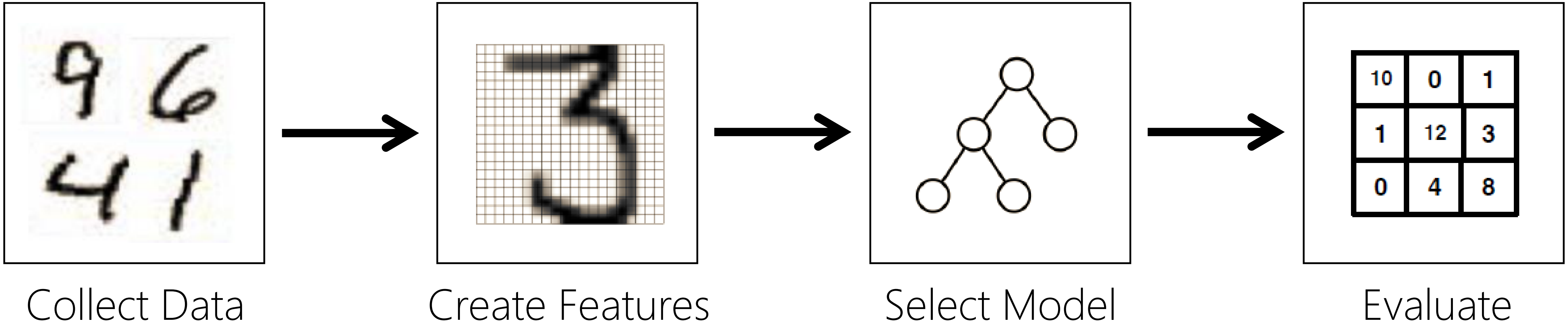
Message #3
From: toni.graham@enron.com
To: lisa.csikos@enron.com, rita.wynne@enron.com, daren.farmer@enron.com
CC: renda.herod@enron.com
Subject: confirming requisitions

Confirming the open requisitions for your group. If your records indicate otherwise, please let me know.

Lisa Csikos 104355, 104001
Rita Wynne 104354
Daren Farmer 104210
Mike Eiben 104323
Pat Clynes 104285

The posting dates have all been updated to reflect a current posting date.
Thanks for your support!!
Toni

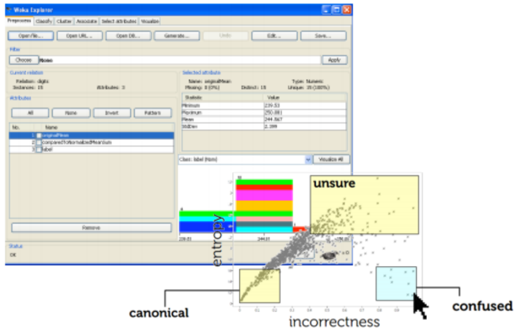
[Stumpf et al, IUI 2007]

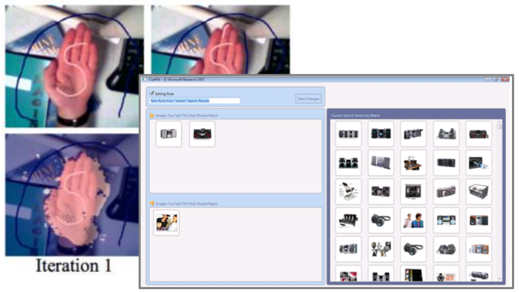

Collect Data
Create Features
Select Model
Evaluate
10 0 1
1 12 3
0 4 8

Experts

Practitioners

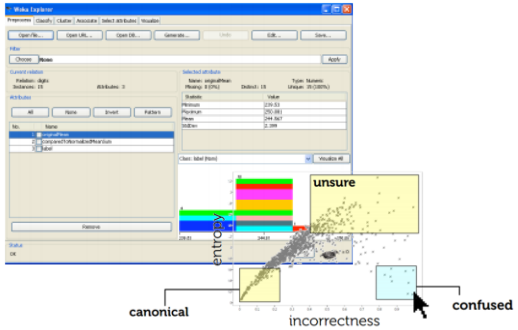
Everyday People



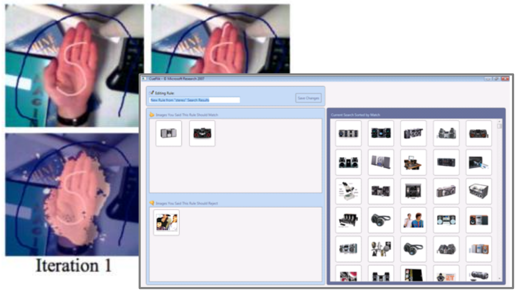


How are these scenarios different?

Experts



Practitioners



Everyday People

User experience impacts what you can expose.

Interaction focus impacts attention and feedback.

Accuracy requirements impacts expected time and effort.

.....

Principles for human-centered ML?

Traditional User Interfaces

Visibility and feedback

Consistency and standards

Predictability

Actionability

Error prevention and recovery

….

Intelligent/ML-Based Interfaces

Safety

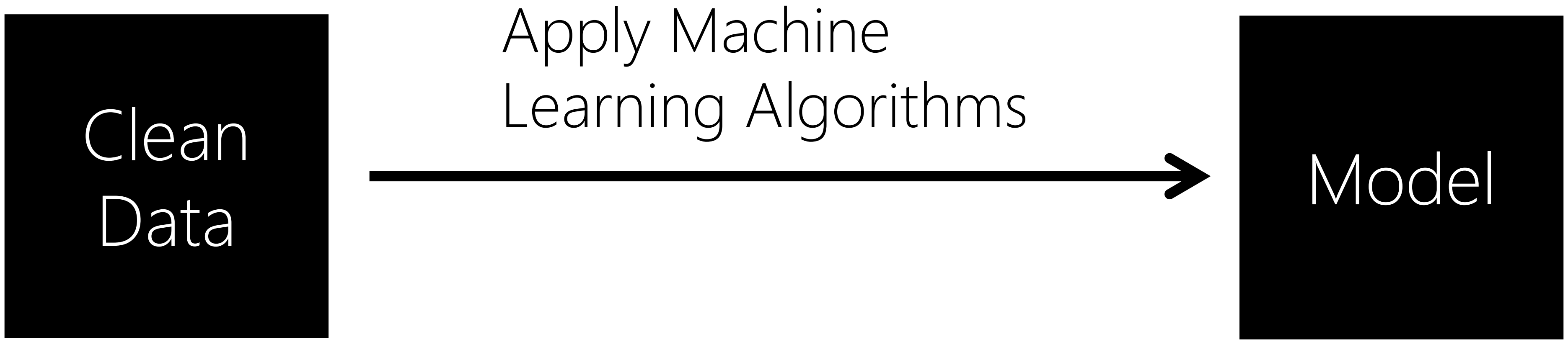
Trust

Manage expectations

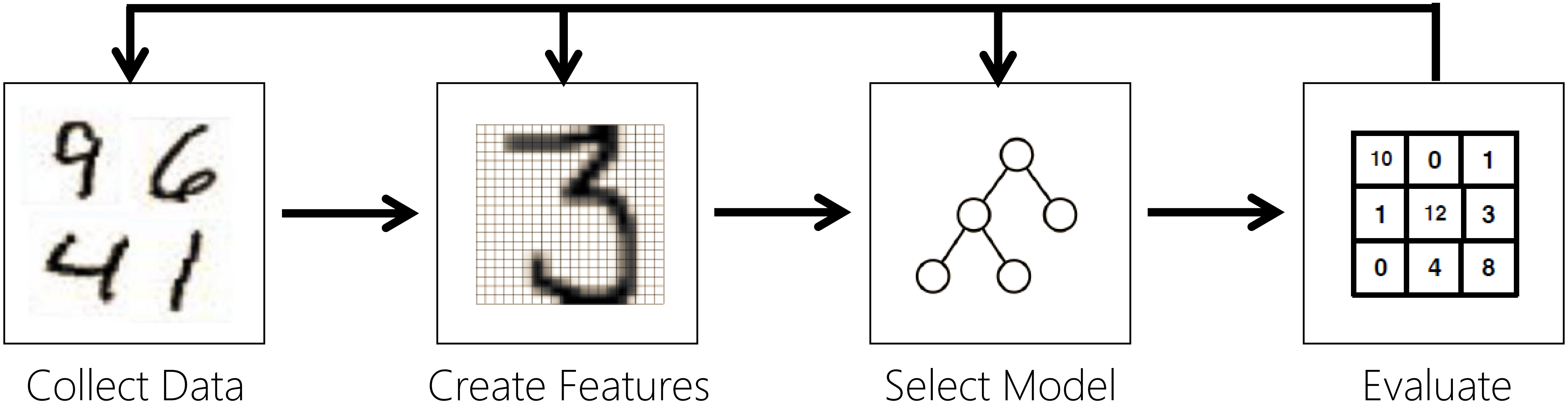
Degrade gracefully under uncertainty

….

Traditional Machine Learning



Human-Centered Machine Learning



Human-Centered Machine Learning

Machine Learning

+

Machine Teaching

samershi@microsoft.com