

VISUALIZATIONS FOR MACHINE LEARNING

(An Introduction)

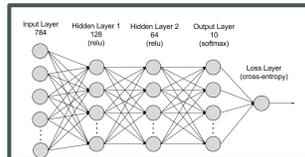
HARI SUBRAMONYAM



<https://xkcd.com/1838/>

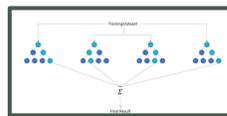
1

What is the zip-code written on the mail? (Handwriting Recognition)



How much will a flight from CA to NYC cost in December? (Flight Price Prediction)

- Airline
- Date of Journey
- Source
- Destination
- Route
- Total Stops
- Departure Time
- Arrival Time
- Pilot Name
- Price
- Duration
- Departure Gate



Machine Learning Pipeline

Liu et al. (2017)

2

In what ways can **visualizations** support machine learning tasks?

3

Understanding Data

What is the quality of the the data?

Is the data representative?

What features are available in the dataset?

Is the test set representative of the data set as a whole?

Is the data labeled correctly?

Model Training

Why is the model behaving the way it is?

What is the structure of the model?

How can I refine the model?

Is this the best technique for modeling the learning task?

How does the model output change with changes in model parameters?

Model Evaluation

How accurate is the model output?

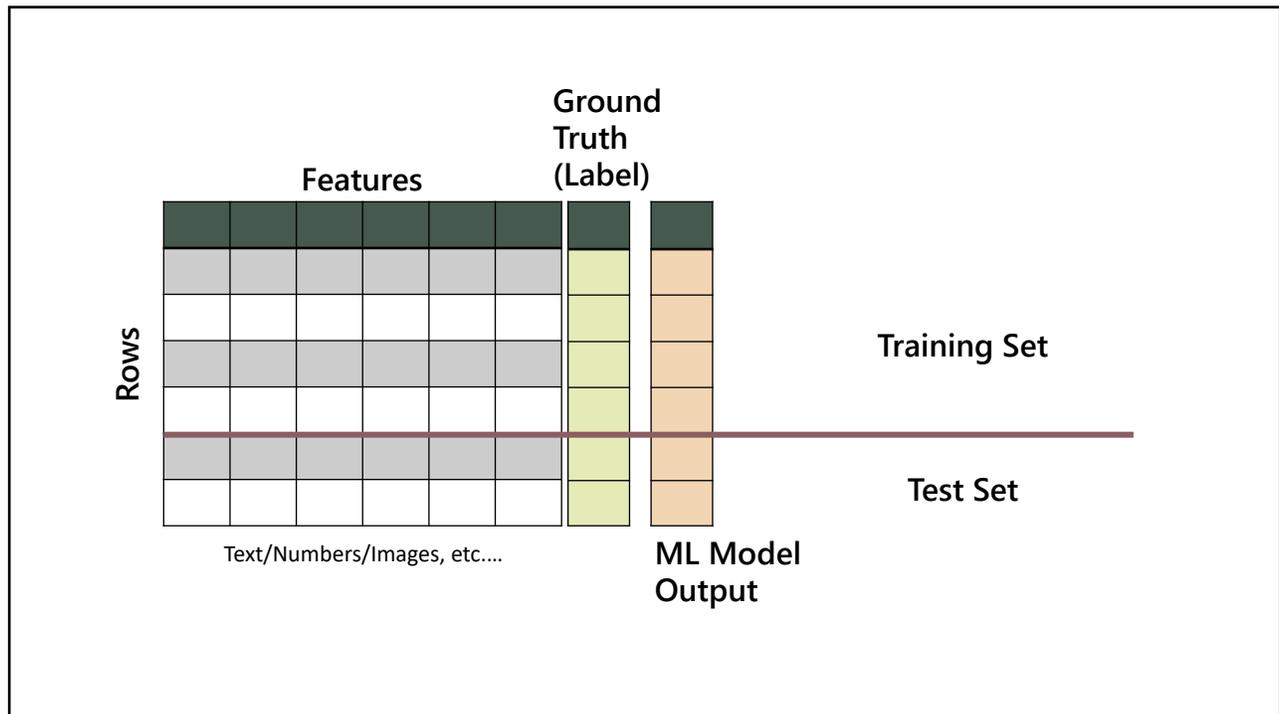
How can I explain the model output?

Is the model fair?

4

Visualizations for Understanding ML Datasets

5



6

- Understanding Data Characteristics before Modeling
- Selecting Features for Modeling
- Debugging Data based on Model Outputs

7

FACETS

The screenshot displays the Facets interface, divided into two main sections: Overview and Dive.

Overview: This section provides a summary of 6 numeric features. It includes a table with columns for count, missing, mean, std dev, zeros, min, median, and max. For each feature, there are two rows of data (32.6k and 16.3k) and a small histogram. The features shown are Age, Capital Gain, Capital Loss, and Education-Num.

Feature	32.6k	16.3k
Age	38.58	38.77
Capital Gain	1,077.65	1,081.91
Capital Loss	402.96	403.11
Education-Num	10.08	10.07

Dive: This section shows a detailed scatter plot of the data. The plot is faceted by feature, with each facet showing a distribution of data points. The plot is color-coded by Occupation, with a legend on the right side. The legend includes categories like Prof-speciality and Exec-managerial.

<https://pair-code.github.io/facets/>

8

- 1994 Census dataset
- ~50k rows
- 14 Attributes (Categorical and Integer)

Prediction task is to determine whether a person makes over 50K a year.

<http://archive.ics.uci.edu/ml/datasets/Census+Income>

Sort by
Non-uniformity Reverse order

Feature search (regex enabled)

Features: int(6) string(9)

■ train ■ test

Chart to show
Standard

log expand percentages

Numeric Features (6)								
	count	missing	mean	std dev	zeros	min	median	max
Capital Gain	32.6k	0%	1,077.65	7,385.29	91.67%	0	0	100k
	16.3k	0%	1,081.91	7,583.94	91.87%	0	0	100k
Capital Loss	32.6k	0%	87.3	402.96	95.33%	0	0	4,356
	16.3k	0%	87.9	403.11	95.31%	0	0	3,770
fnlwgt	32.6k	0%	190k	106k	0%	12.3k	178k	1.48M
	16.3k	0%	189k	106k	0%	13.5k	178k	1.49M
Hours per week	32.6k	0%	40.44	12.35	0%	1	40	99
	16.3k	0%	40.39	12.48	0%	1	40	99
Education-Num	32.6k	0%	10.08	2.57	0%	1	10	16
	16.3k	0%	10.07	2.57	0%	1	10	16
Age	32.6k	0%	38.58	13.64	0%	17	37	90
	16.3k	0%	38.77	13.85	0%	17	37	90

9

- 1994 Census dataset
- ~50k rows
- 14 Attributes (Categorical and Integer)

Prediction task is to determine whether a person makes over 50K a year.

<http://archive.ics.uci.edu/ml/datasets/Census+Income>

Categorical Features (9)

	count	missing	unique	top	freq top	avg str len
Target	32.6k	0%	2	<=50K	24.7k	4.76
	16.3k	0%	2	<=50K.	12.4k	5.76

SHOW RAW DATA

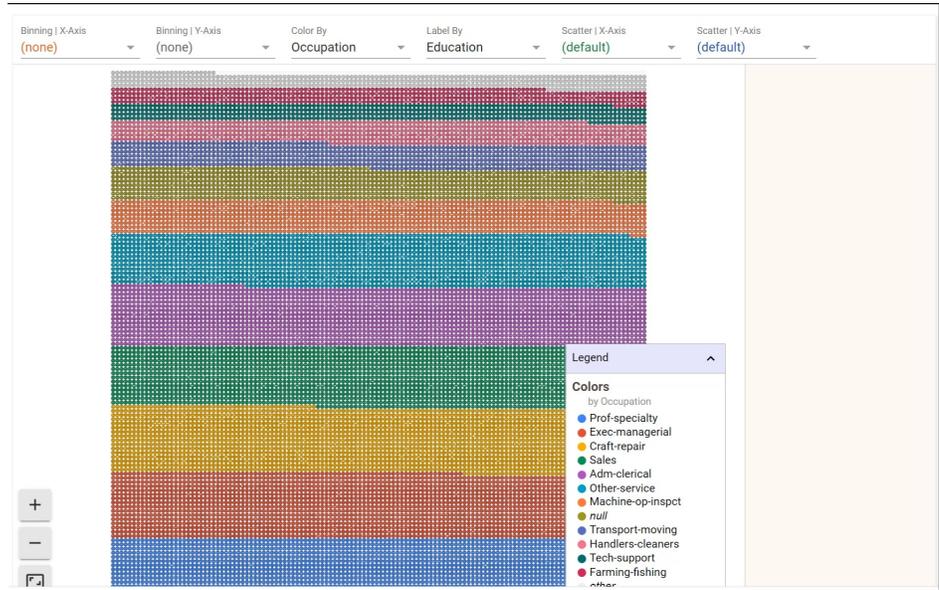
Chart to show
Standard

log expand percentages

10

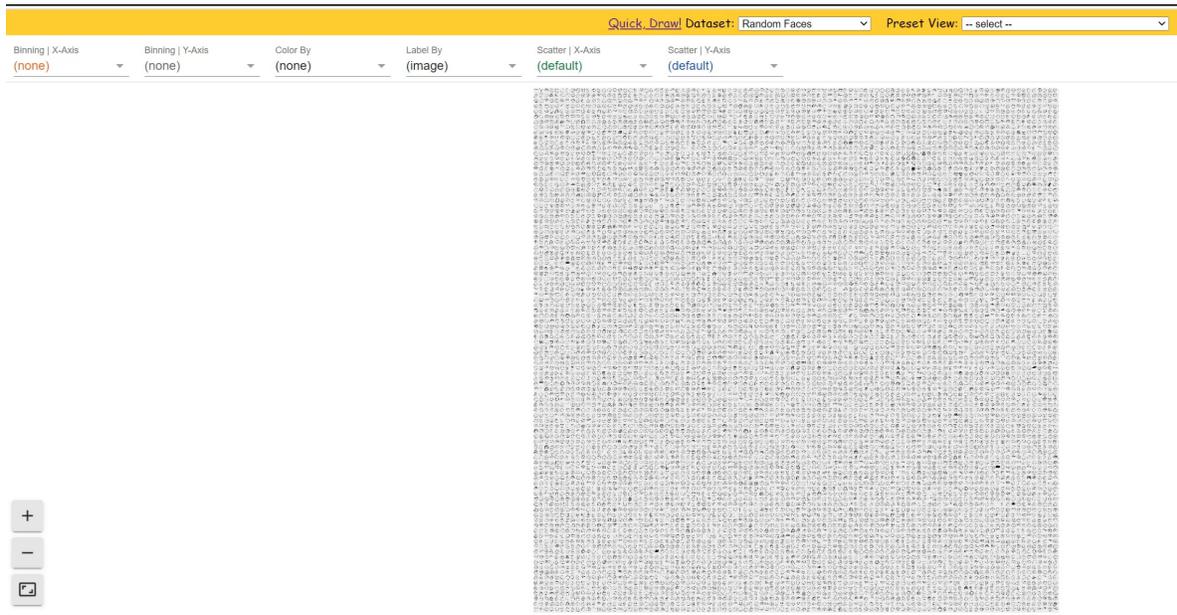
- 1994 Census dataset
- ~50k rows
- 14 Attributes (Categorical and Integer)

Prediction task is to determine whether a person makes over 50K a year.



<http://archive.ics.uci.edu/ml/datasets/Census+Income>

11



<https://pair-code.github.io/facets/quickdraw.html>

12

Randomly sampled sketches (left), pixel overlay (center) and sketch-rnn average (right) of telephone drawings from around the world.

Randomly sampled sketches (left), pixel overlay (center) and sketch-rnn average (right) of telephone drawings from India only.

Japan

United States

Start
Counter clockwise
Clockwise

<https://qz.com/994486/the-way-you-draw-circles-says-a-lot-about-you/>
<https://medium.com/analytics-vidhya/analyzing-sketches-around-the-world-with-sketch-rnn-c6cbe9b5ac80>

13

INFUSE (INteractive FeatURE SElection)

Sort by name Sort by selection

g1

- GLOM FILT RATE, EST
- GLUCOCORTICOIDS
- GLUCOSE
- GLUCOSE - 1 HOUR
- GLUCOSE - 2 HOUR
- GLUCOSE - 3 HOUR
- GLUCOSE, 1 HOUR
- GLUCOSE, 1/2 HOUR
- GLUCOSE, 2 HOUR
- GLUCOSE, 3 HOUR
- GLUCOSE, TWO-HOUR POSTPRANDIAL
- HEMOGLOBIN A1C/HEMOGLOBIN.TOTA
- HYPERGLYCEMICS
- ANTHYPERGLY, INCRETIN MIMETIC/GLP-1 RECEPTOR AGONIST
- ANTHYPERGLY, (DPP-4) INHIBITOR & BIGUANIDE COMB.
- ANTHYPERGLYCEMIC, AMYLIN ANALOG-TYPE
- ANTHYPERGLYCEMIC, DPP-4 INHIBITORS
- ANTHYPERGLYCEMIC, INSULIN-RELEASE STIMULANT TYPE
- ANTHYPERGLYCEMIC, INSULIN-RESPONSE ENHANCER (N-S)
- ANTHYPERGLYCEMIC, BIGUANIDE TYPE (NON-SULFONYLUREA)
- ANTHYPERGLYCEMIC, INSULIN-REL STIM. & BIGUANIDE CMB
- ANTHYPERGLYCEMIC, INSULIN-RESP ENHANCER & BIGUANIDE CMB
- DIGITALIS GLYCOSIDES
- Prim Open Angle Glaucoma
- Hypoglycemia NOS
- Abn Glucose Toler-Deliv
- Abn Glucose-Antepartum
- Pure Hyperglycemia
- Abnormal Glucose NEC

	Tree	LogisticRegression	NaiveBayes	SVM
S0	6383	6108	6203	4584
S1	6309	6110	6124	4693
S2	5721	5752	5805	5048
S3	5721	5748	5798	5050

DIAGNOSIS - HCC
 DIAGNOSIS - ProblemList
 Lab - Lab
 Medication - Ingredient
 Medication - Orders
 PROCEDURE - Cpt code

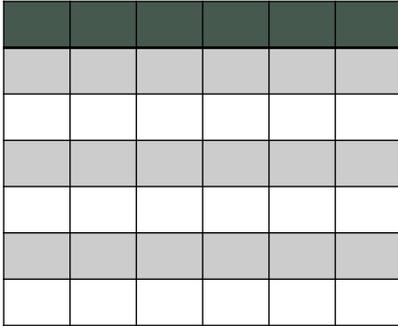
Switch Layout Average Rank Pick Count Inward Pie Glyph Reset View Evaluate Model

Krause, Perer, and Bertini (2014)

14

Domain Task: Predict if a patient is at risk of developing diabetes.

858 Features



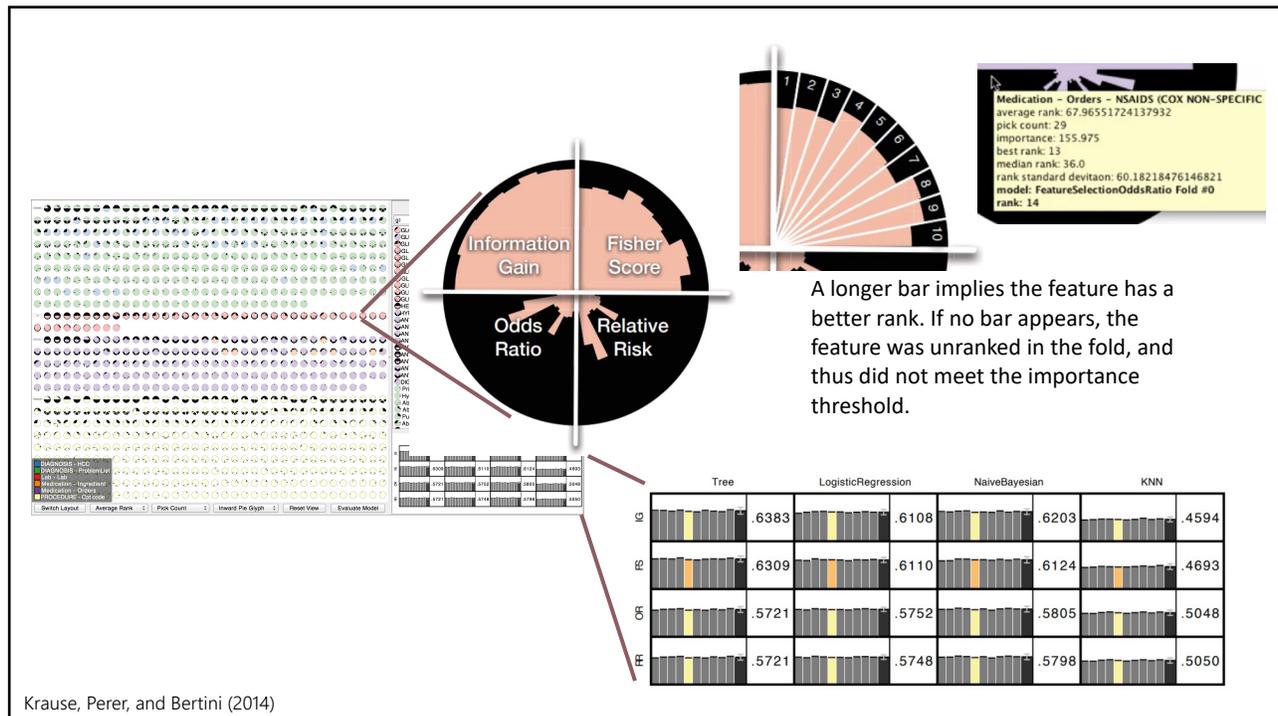
Diabetes Dataset

ML Task 1: Comparison of feature selection algorithms. (4 Algorithms)

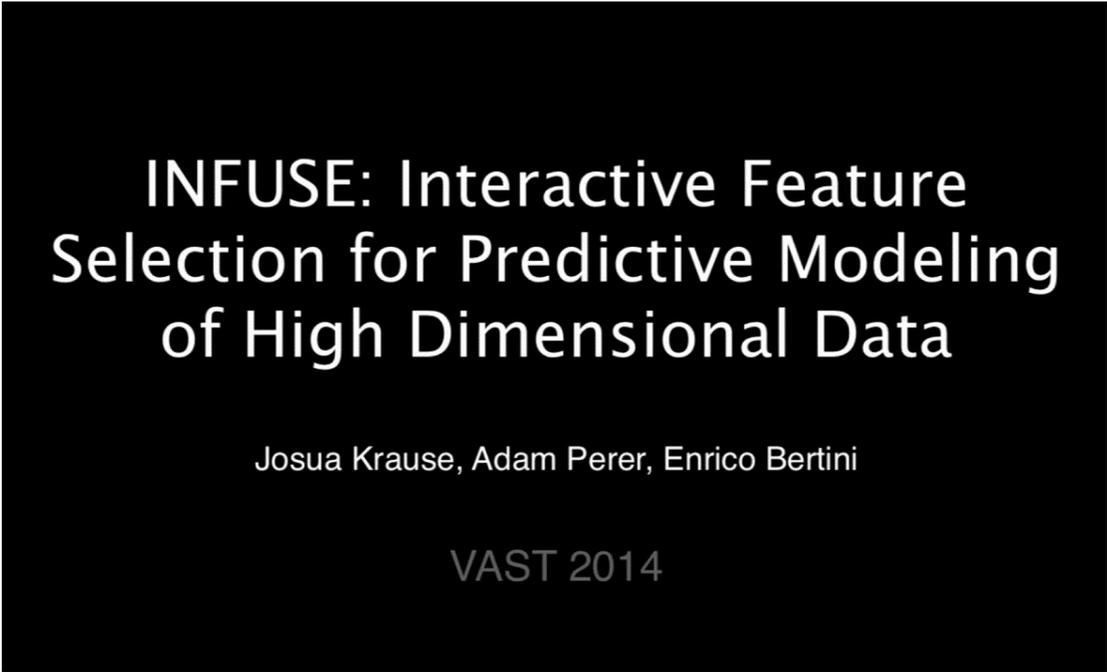
ML Task 2: Comparison of classification algorithms. (4 classifiers)

ML Task 3: Manual selection and testing of feature sets

15



16



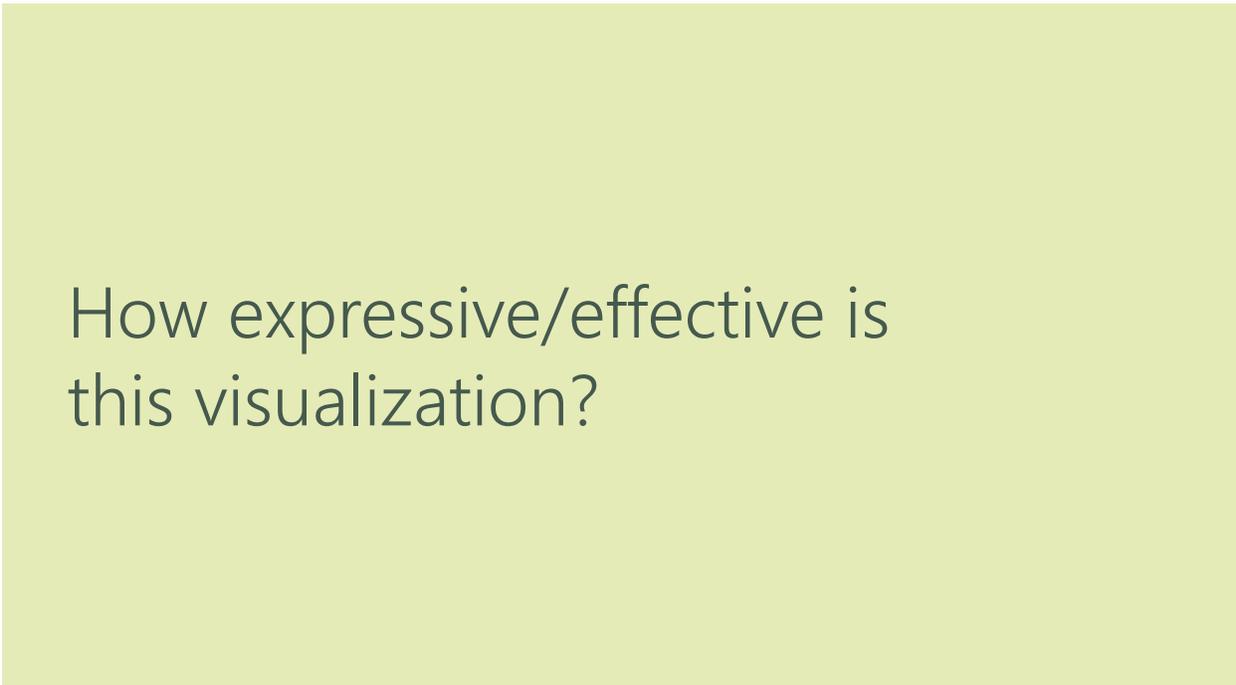
INFUSE: Interactive Feature Selection for Predictive Modeling of High Dimensional Data

Josua Krause, Adam Perer, Enrico Bertini

VAST 2014

Krause, Perer, and Bertini (2014)

17



How expressive/effective is this visualization?

18

DataDebugger

(a) (b) (c) (d) (e)

Xiang et al. (2019)

19

Web Images

amazon ebay Taobao

- Clothes Image Dataset
- 37,000 Instances
- 14 Categories

ML Task: Classify article of clothing based on image

T-shirt

Jacket

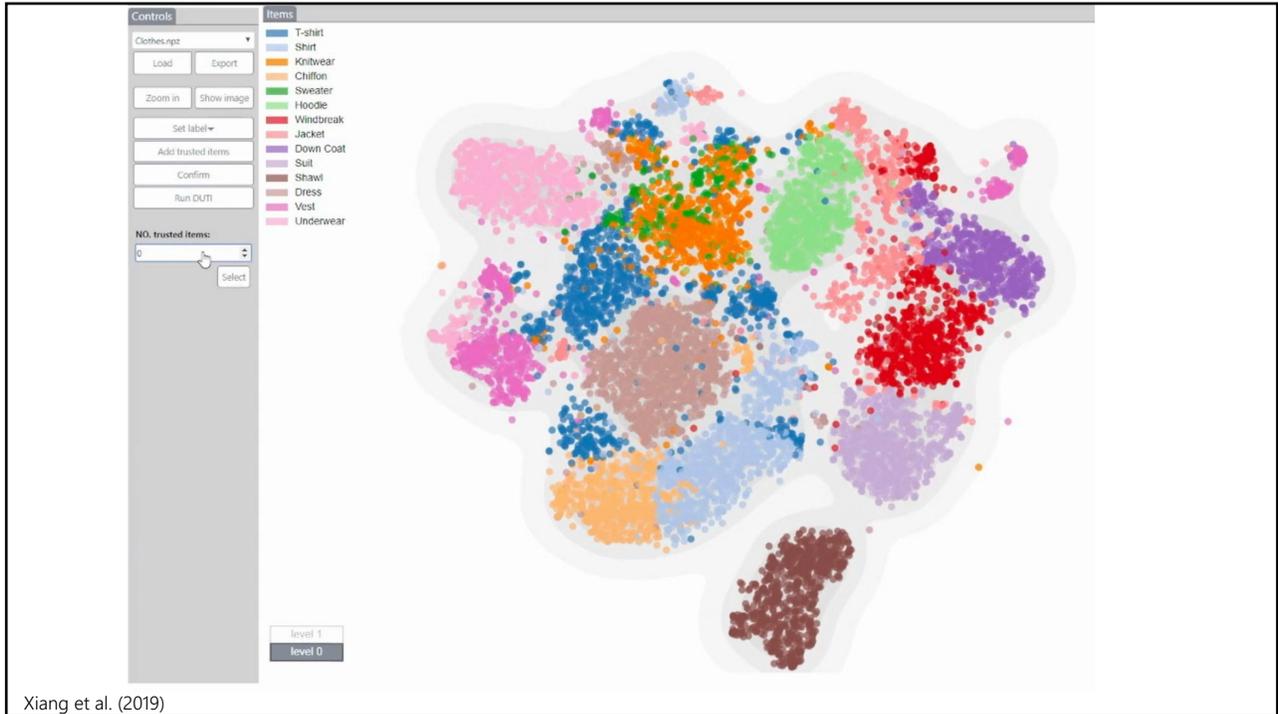
T-shirt

Jacket

Jacket

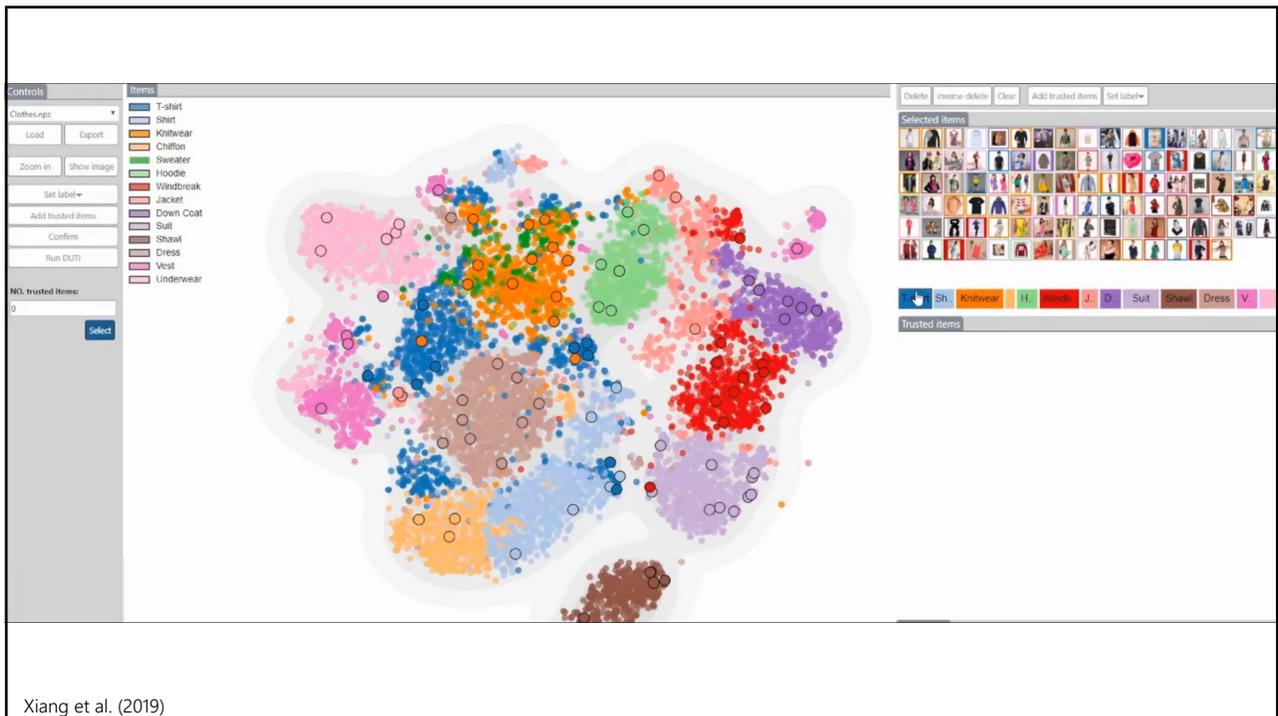
Xiang et al. (2019)

20



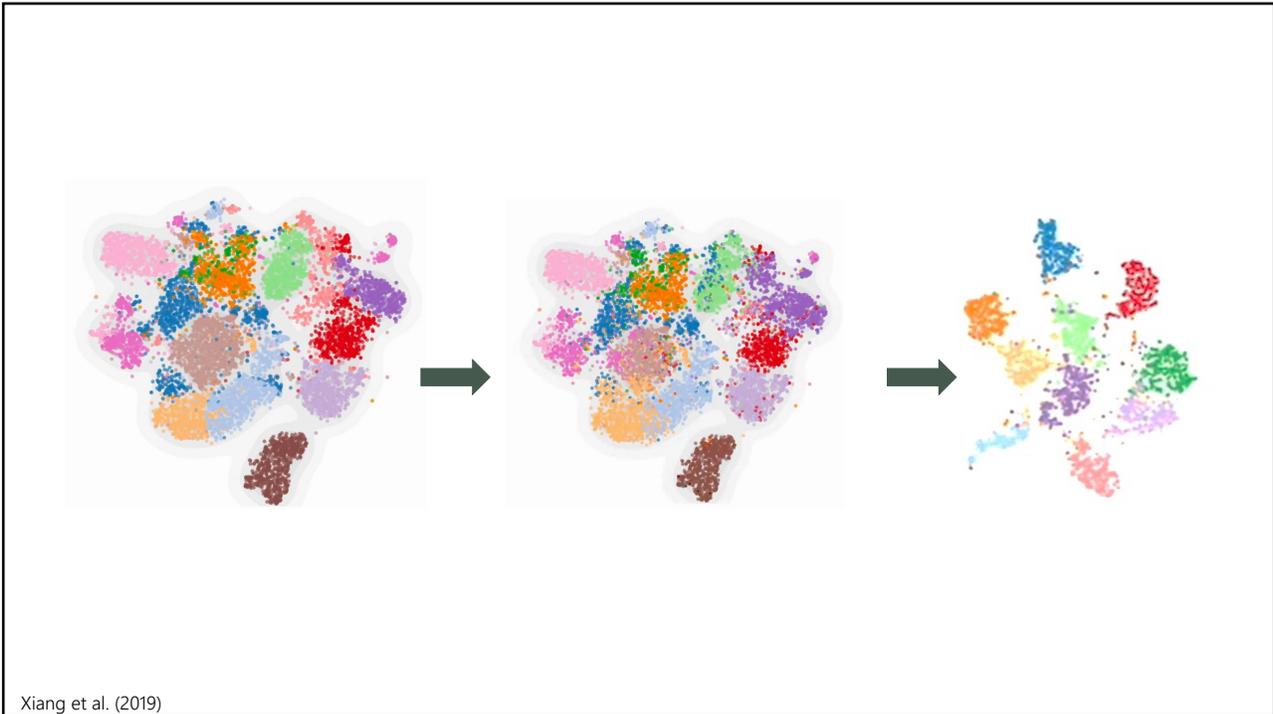
Xiang et al. (2019)

23



Xiang et al. (2019)

24



25

Errudite

Error Analysis by:

- Expressive grouping of error instance
- Counterfactual evaluation

Errudite: An Interactive Tool for Scalable and Reproducible Error Analysis

Wu et al. (2019)

26

Find instances to explore (without edits)

Get Instances | Sample 10 instances **worst predicted by bidaf** that are in | Select groups | and not in | Select groups

Filter CMD | Preview the filter on 10570 instances

Filtered instances: 10570 (100.0% of total), Error: 3378 (32.0% of slice, 32.0% of total, 100.0% of all errors)

Record the Group | Get samples

Returned instances (answer encoding: groundtruth, prediction by bidaf (correct, incorrect), model prediction distributions)

What year did Börte's give birth to Jochi?

As previously arranged by his father, Temüjin married Börte of the Onggirat tribe when he was around 16 in order to cement alliances between their tribes. Soon after Börte's marriage to Temüjin, she was kidnapped by the Merkits and reportedly given away as a wife. Temüjin rescued her with the help of his friend and future rival, Jamukha, and his protector, Toghrul Khan of the Keraité tribe. She gave birth to a son, Jochi (1185-1226), nine months later, clouding the issue of his parentage. Despite speculation over Jochi, Börte would be Temüjin's only empress, though he did follow tradition by taking several morganatic wives.

When did Wei Yilin die?

The Chinese medical tradition of the Yuan had "Four Great Schools" that the Yuan inherited from the Jin dynasty. All four schools were based on the same intellectual foundation, but advocated different theoretical approaches toward medicine. Under the Mongols, the practice of Chinese medicine spread to other parts of the empire. Chinese physicians were brought along military campaigns by the Mongols as they expanded towards the west. Chinese medical techniques such as acupuncture, moxibustion, pulse diagnosis, and various herbal drugs and elixirs were transmitted westward to the Middle East and the rest of the world. Several medical advances were made in the Yuan period. The physician Wei Yilin (1277-1342) invented a suspension method for reducing dislocated joints, which he performed using anesthetics. The Mongol physician Hu Sihui described the importance of a healthy diet in a 1330 medical treatise.

Wu et al. (2019)

Error Instances

27

rudrude: An Interactive Tool for Scalable and Reproducible Error Analysis

Model Comparison

Model	em	f1	sent	precision	recall
bidaf	0.68	0.77	0.91	0.78	0.81

Attributes Overview

- answer_type: groundtruth
- question_type: question
- context_length: length(context)
- question_length: length(question)

Find instances to explore (without edits)

Get Instances | Sample 10 instances **worst predicted by bidaf** that are in | Select groups | and not in | Select groups

Filter CMD | Preview the filter on 10570 instances

Filtered instances: 10570 (100.0% of total), Error: 3378 (32.0% of slice, 32.0% of total, 100.0% of all errors)

Record the Group | Get samples

Returned instances (answer encoding: groundtruth, prediction by bidaf (correct, incorrect), model prediction distributions)

What year did Börte's give birth to Jochi?

As previously arranged by his father, Temüjin married Börte of the Onggirat tribe when he was around 16 in order to cement alliances between their respective tribes. Soon after Börte's marriage to Temüjin, she was kidnapped by the Merkits and reportedly given away as a wife. Temüjin rescued her with the help of his friend and future rival, Jamukha, and his protector, Toghrul Khan of the Keraité tribe. She gave birth to a son, Jochi (1185-1226), nine months later, clouding the issue of his parentage. Despite speculation over Jochi, Börte would be Temüjin's only empress, though he did follow tradition by taking several morganatic wives.

When did Wei Yilin die?

The Chinese medical tradition of the Yuan had "Four Great Schools" that the Yuan inherited from the Jin dynasty. All four schools were based on the same intellectual foundation, but advocated different theoretical approaches toward medicine. Under the Mongols, the practice of Chinese medicine spread to other parts of the empire. Chinese physicians were brought along military campaigns by the Mongols as they expanded towards the west. Chinese medical techniques such as acupuncture, moxibustion, pulse diagnosis, and various herbal drugs and elixirs were transmitted westward to the Middle East and the rest of the world. Several medical advances were made in the Yuan period. The physician Wei Yilin (1277-1342) invented a suspension method for reducing dislocated joints, which he performed using anesthetics. The Mongol physician Hu Sihui described the importance of a healthy diet in a 1330 medical treatise.

Wu et al. (2019)

28

Errudite: An Interactive Tool for Scalable and Reproducible Error Analysis

Metrics Overview

Model	em	f1	sent	precision	recall
Model	0.68	0.77	0.91	0.78	0.81
bidaf					

Attributes

- answer_type: `answer_type(groundtruth)`
- question_type: `question_type(question)`
- context_length: `length(context)`
- question_length: `length(question)`
- groundtruths_length: `length(groundtruths)`

Find instances to explore

Get Instances: Sample 10 instances worst predicted by bidaf - that are in [Select groups] and not in [Select groups]

Filter CMD: `STRING(groundtruth) in STRING(prediction (model="ANCHOR")) and F1 (model="ANCHOR") == 0`

Filtered instances: 26 (0.2% of total), Error: 26 (100.0% of slice, 0.2% of total, 0.8% of all errors)

Returned instances (answer encoding: groundtruth_prediction by bidaf(correct_incorrect)_model prediction distributions)

John Mayow died in what year?

In the late 17th century, Robert Boyle proved that air is necessary for combustion. English chemist John Mayow (1641-1679) refined this work by showing that fire requires only a part of air that he called spiritus nitroaereus or just nitroaereus. In one experiment he found that placing either a mouse or a lit candle in a closed container over water caused the water to rise and replace one-fourteenth of the air's volume before extinguishing the subjects. From this he surmised that nitroaereus is consumed in both respiration and combustion.

What year did Börte's give birth to Jochi?

As previously arranged by his father, Temüjin married Börte of the Onggirat tribe when he was around 16 in order to cement alliances between their respective tribes. Soon after Börte's marriage to Temüjin, she was kidnapped by the Merkits and reportedly given away as a wife. Temüjin rescued her with the help of his friend and future rival, Jamukha, and his protector, Toghrul Khan of the Keraites tribe. She gave birth to a son, Jochi (1185-1228), nine months later, clouding the issue of his parentage. Despite speculation over Jochi, Börte would be Temüjin's only empress, though he did follow tradition by taking several morganatic wives.

When did Wei Yilin die?

Displaying #0-10 samples.

Wu et al. (2019)

29

Errudite: An Interactive Tool for Scalable and Reproducible Error Analysis

Metrics Overview

Model	accuracy
saa	0.62
vgcounting	0.66

Attributes

- answer_type: `answer_type(groundtruth)`
- question_type: `question_type(question)`
- question_length: `length(question)`
- groundtruths_length: `length(groundtruths)`

Find instances to explore

Get Instances: Sample 10 instances randomly - that are in [Select groups] and not in [Select groups]

Filter CMD: `attr:count_g > 5`

Filtered instances: 16490 (13.6% of total), Error: 14098 (85.5% of slice, 11.6% of total, 29.9% of all errors)

Returned instances (Correct prediction, incorrect prediction)

What color is the sheets?

groundtruth: striped (* 3) | groundtruth: white (* 2) | groundtruth: stripes (* 1) | groundtruth: white striped (* 1) | groundtruth: multicolored (* 1) | groundtruth: white with blue, yellow, and red stripes (* 1) | groundtruth: multi-color (* 1)

Model predictions: `saa: blue` | `vgcounting: white`

What is on the sandwiches?

groundtruth: meat, cheese, lettuce (* 1) | groundtruth: tomato and lettuce (* 1) | groundtruth: lettuce, cheese and tomato (* 1) | groundtruth: vegetables (* 1) | groundtruth: ham, lettuce, tomatoes (* 1) | groundtruth: lettuce, tomatoes, cabbage (* 1) | groundtruth: lettuce meat tomato (* 1) | groundtruth: bit (* 1) | groundtruth: ham and lettuce (* 1) | groundtruth: avocado (* 1)

Model predictions: `saa: lettuce` | `vgcounting: eaters`

Displaying #0-10 samples.

Wu et al. (2019)

30

Programming by Demonstration

American composer [John Debney](#) created a new arrangement of Ron Grainer's original theme for Doctor Who in 1996.

For the return of the series in 2005, **Murray Gold** provided a new arrangement which featured samples from the 1963 original with further elements added; in the 2005 Christmas episode "The Christmas Invasion", Gold introduced a modified closing credits arrangement that was used up until the conclusion of the 2007 series. [citation needed]

DID YOU MEAN TO FILTER INSTANCES THAT ARE...

```

④ starts_with(prediction(model="bidaf"), pattern="NNP")
④ starts_with(prediction(model="bidaf"), pattern="PERSON")
④ attr:answer_type == answer_type(prediction(model="bidaf"))
④ exact_match(model="bidaf") == 0
④ is_correct_sent(prediction(model="bidaf")) == 0
④ overlap(question, sentence(prediction(model="bidaf"))) > overlap(question, sentence(groundtruths))

```

Wu et al. (2019)

31

How does this approach compare to conventional GUI input elements ?

32

Visualizations for Modeling

33

RuleMatrix



Ming, Qu, & Bertini (2019)

34

Domain Questions

What knowledge has the model learned?

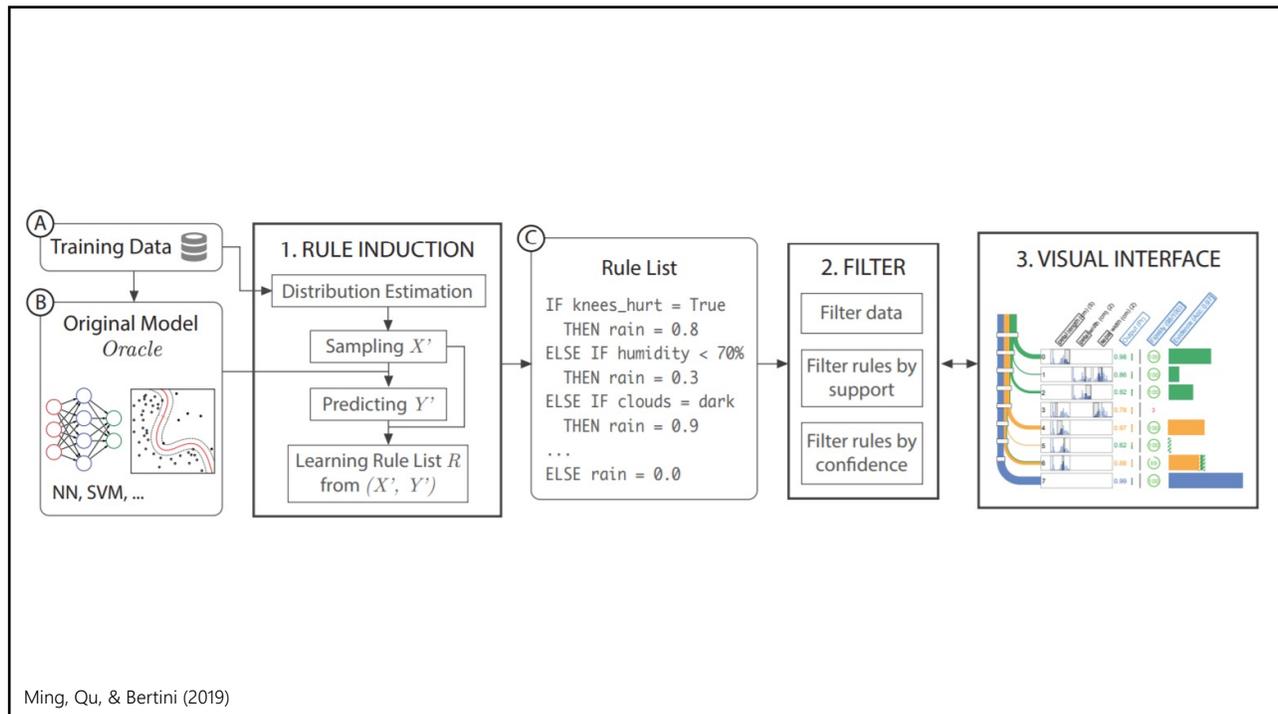
How certain is the model for each piece of knowledge?

What knowledge does the model utilize to make a prediction?

When and where is the model likely to fail?

Ming, Qu, & Bertini (2019)

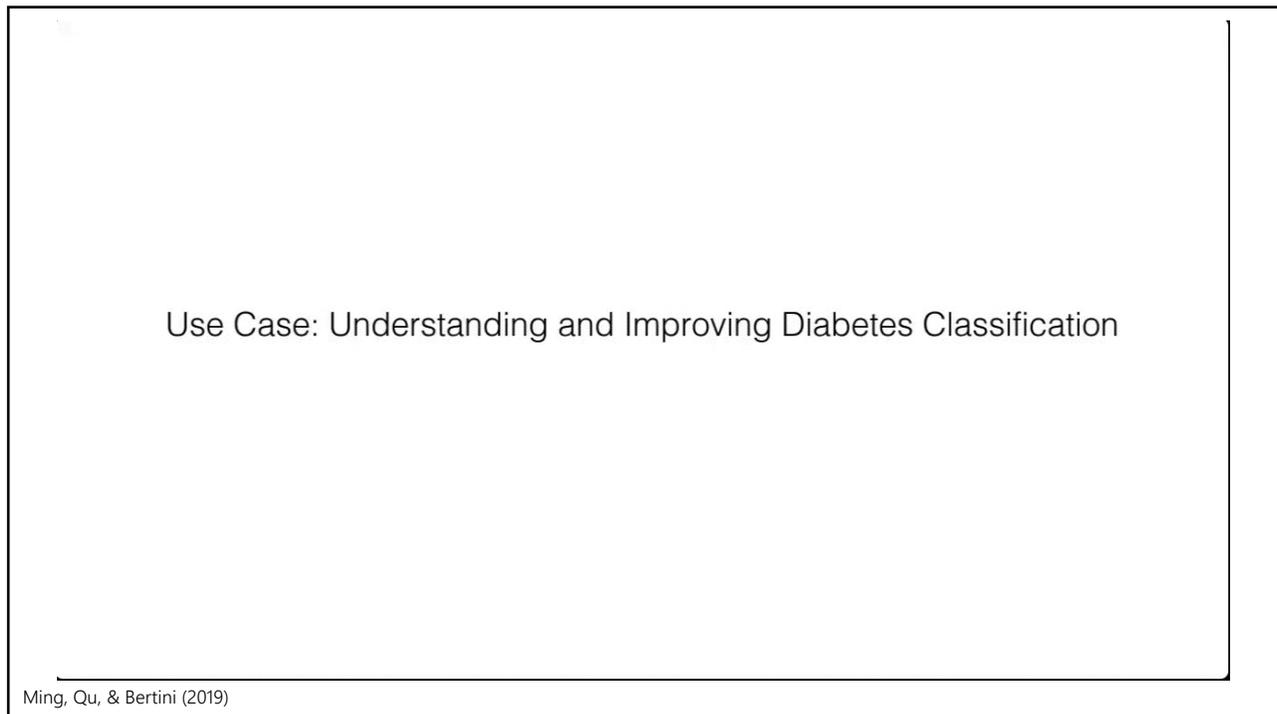
35



36

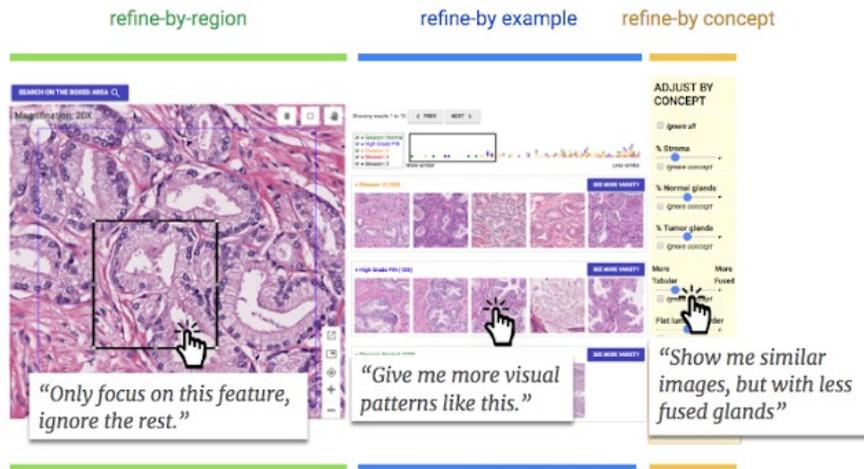


37



38

SMILY (Similar Medical Images Like Yours)



Cai et al. (2019)

39

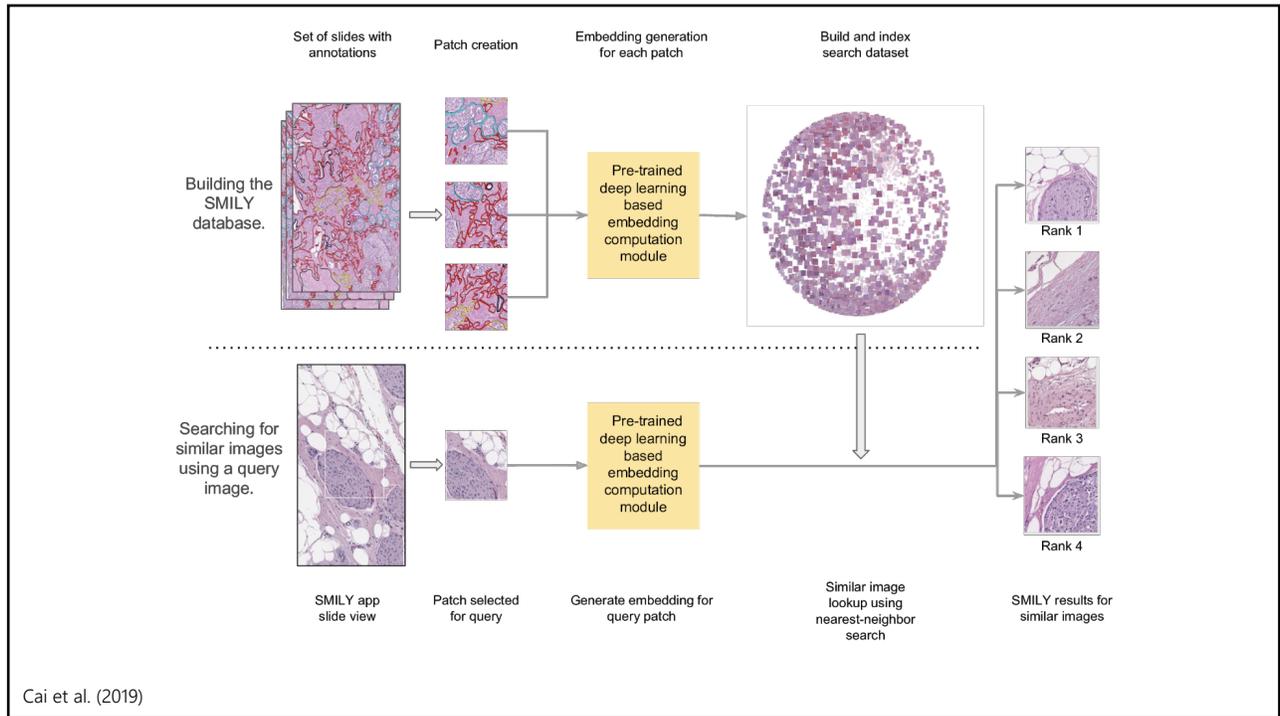
Domain Task

Pathologists need to retrieve visually similar medical images from past patients (e.g. tissue from biopsies) to reference when making a medical decision with a new patient.

Control which types of similarity matter

Cai et al. (2019)

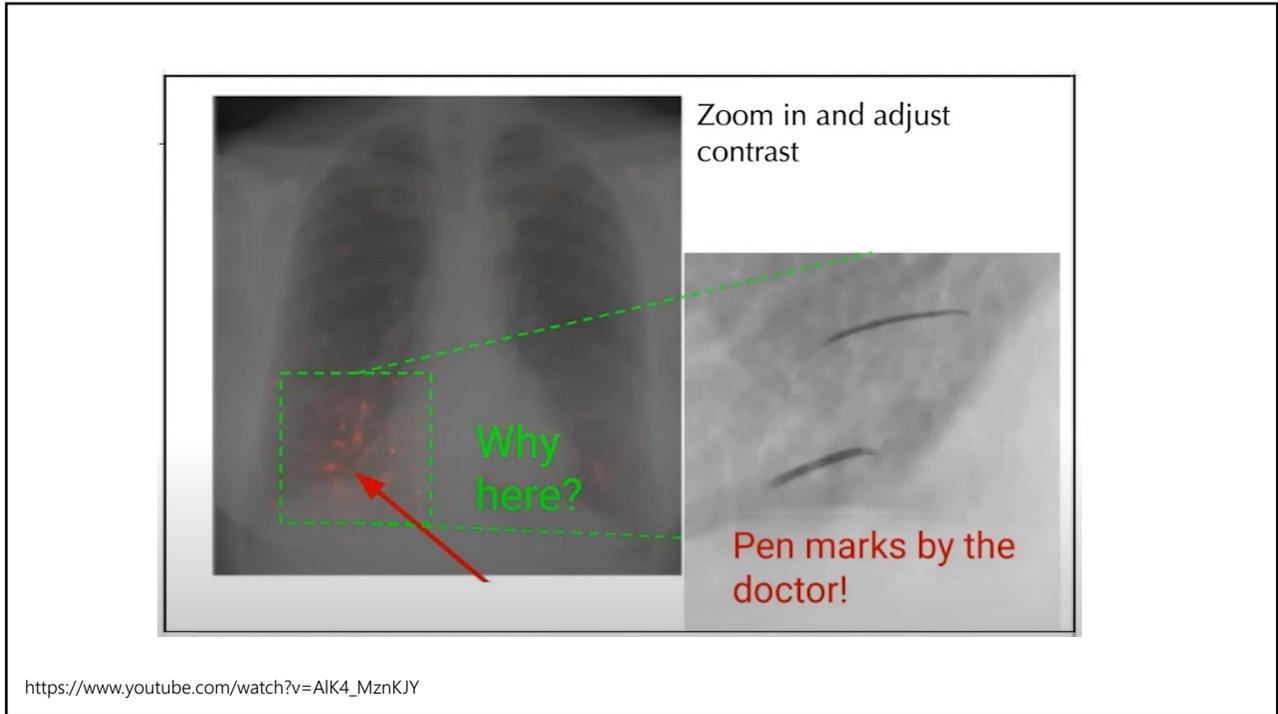
40



41



42

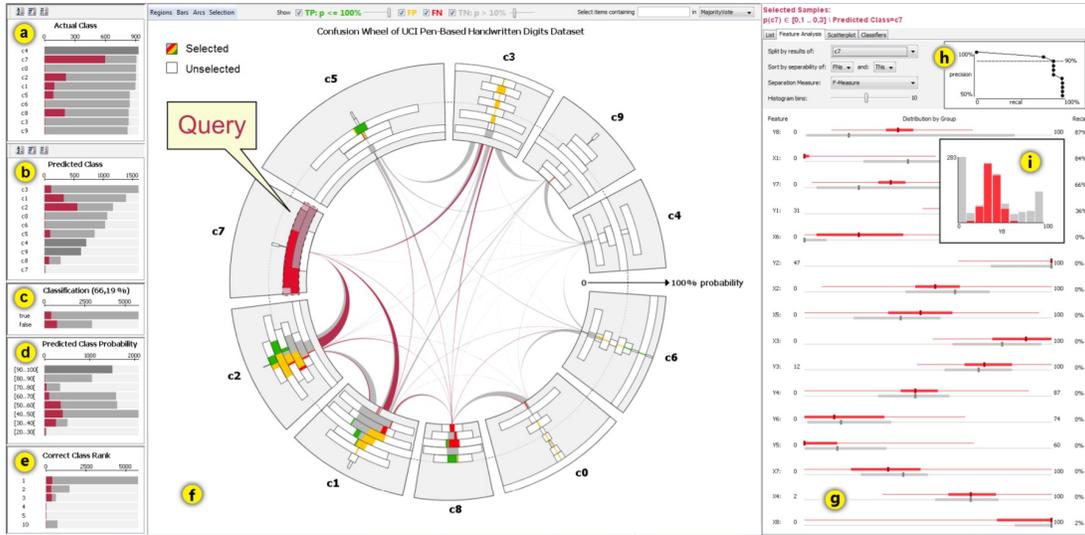


43

Visualizations for Model Evaluation

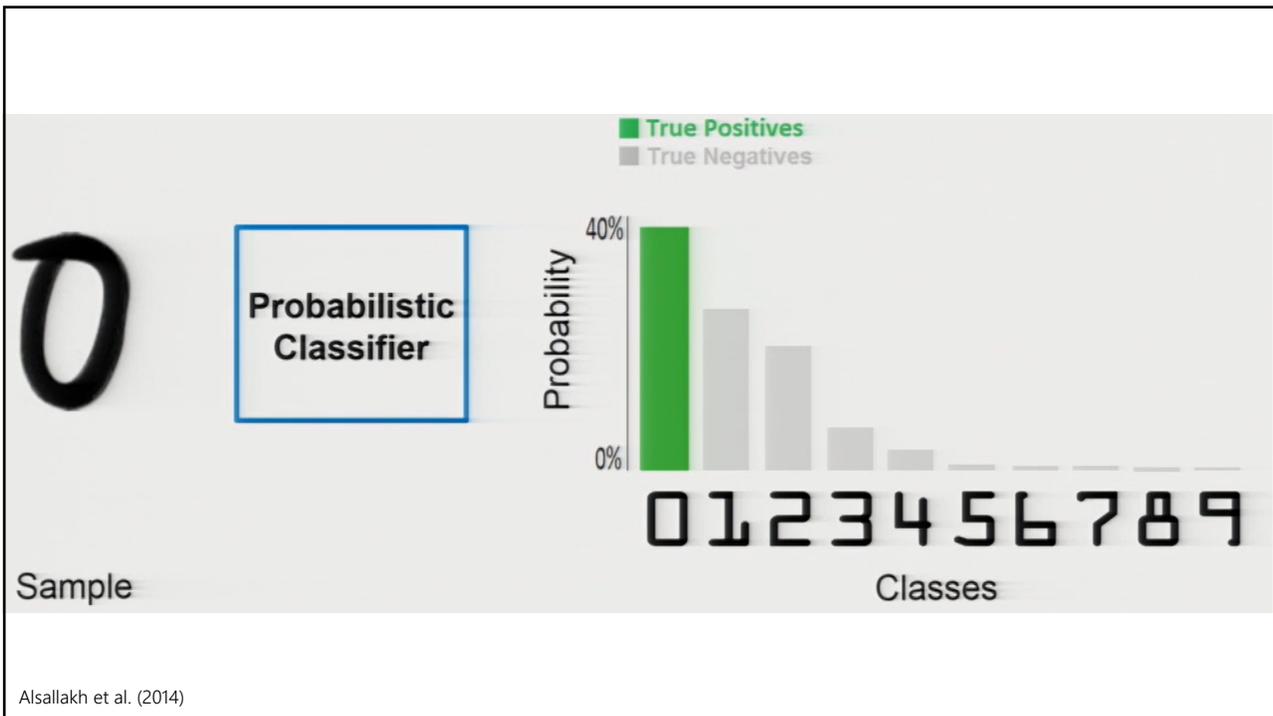
44

Confusion Wheel



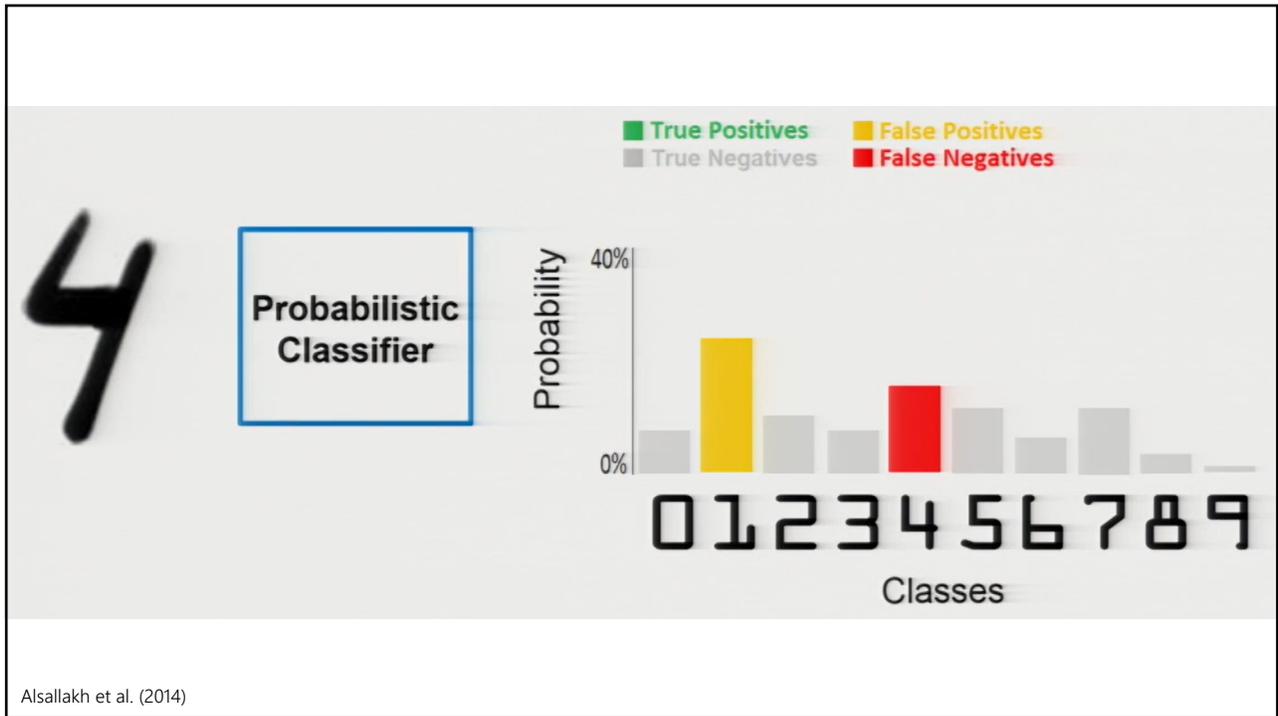
Alsallakh et al. (2014)

45

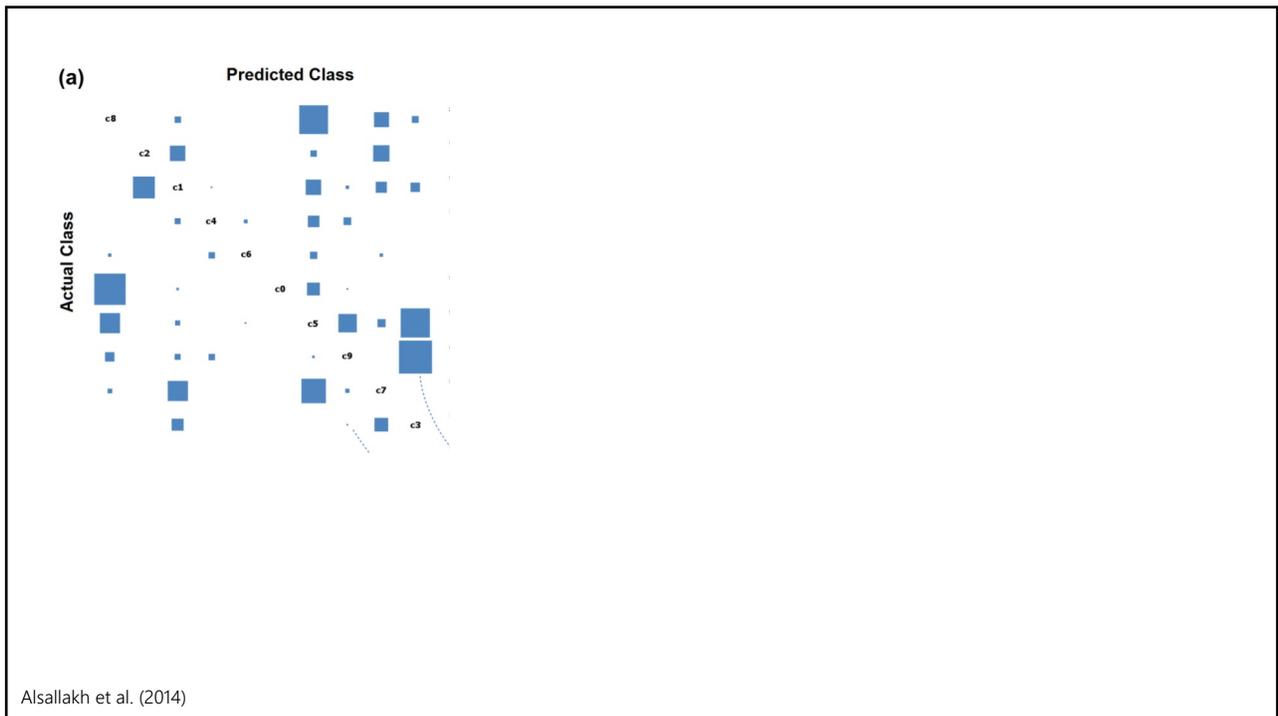


Alsallakh et al. (2014)

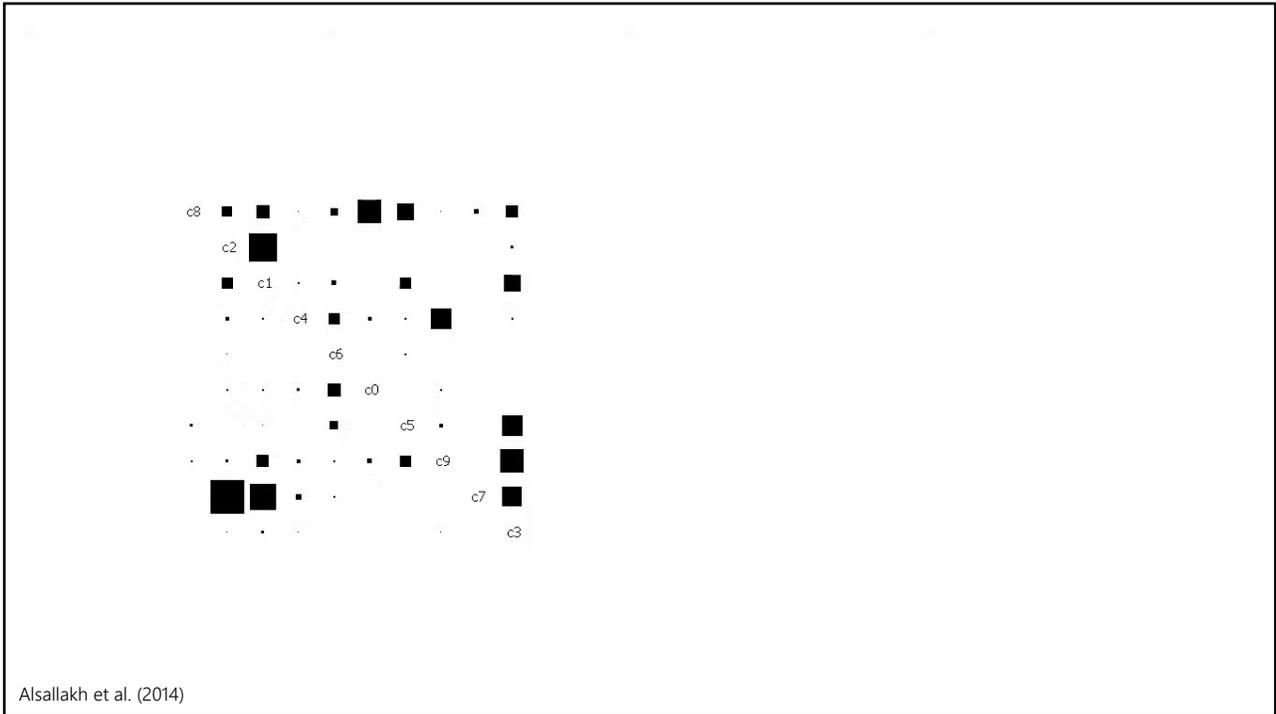
46



47



48

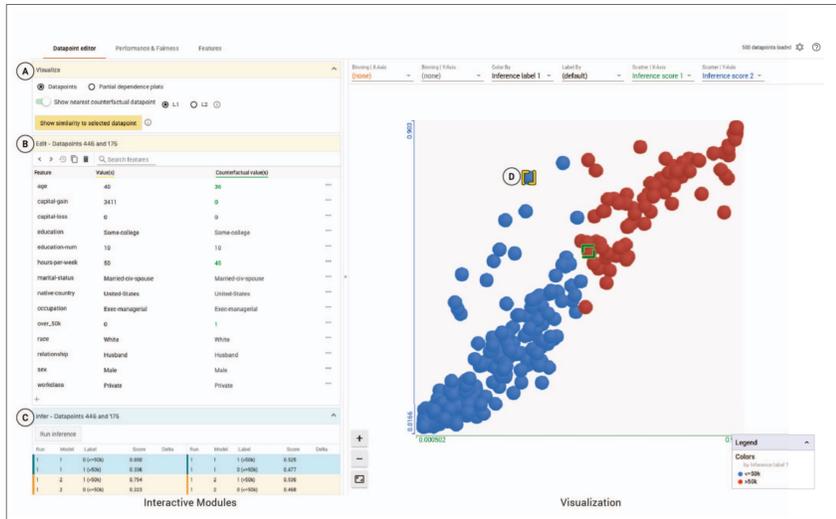


49

In what other ways can we visualize this data?

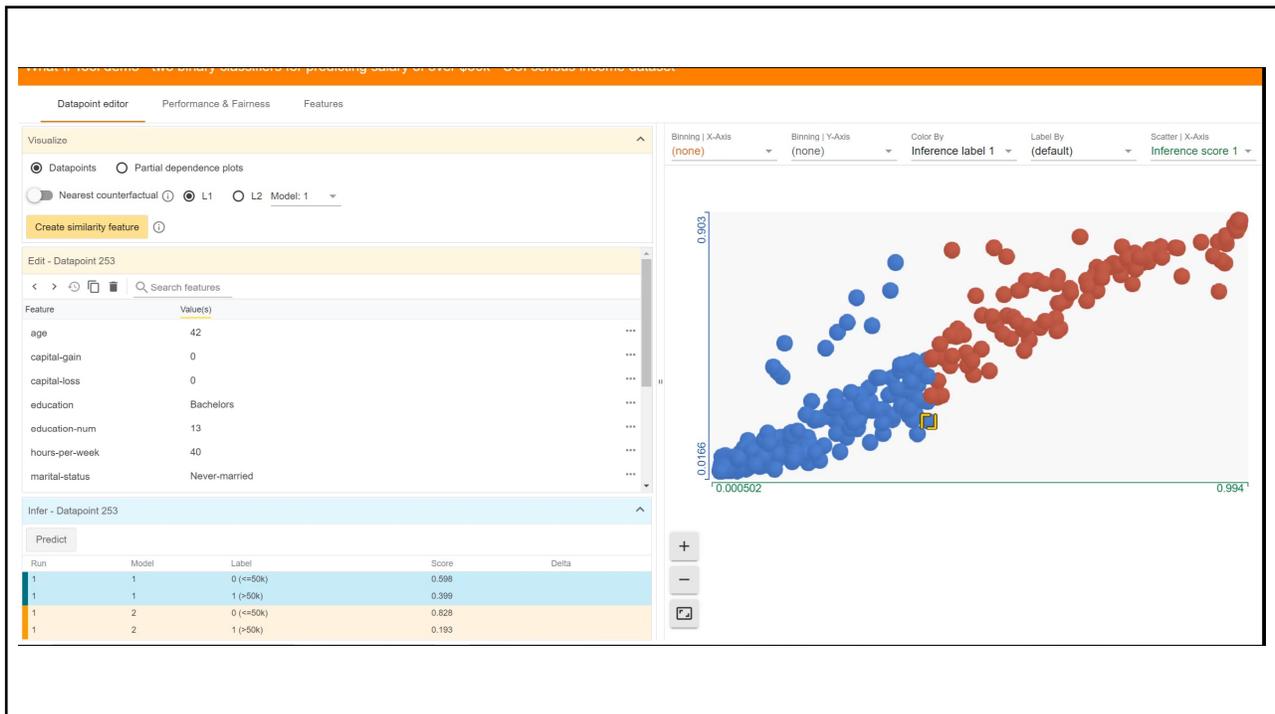
50

The What-If Tool

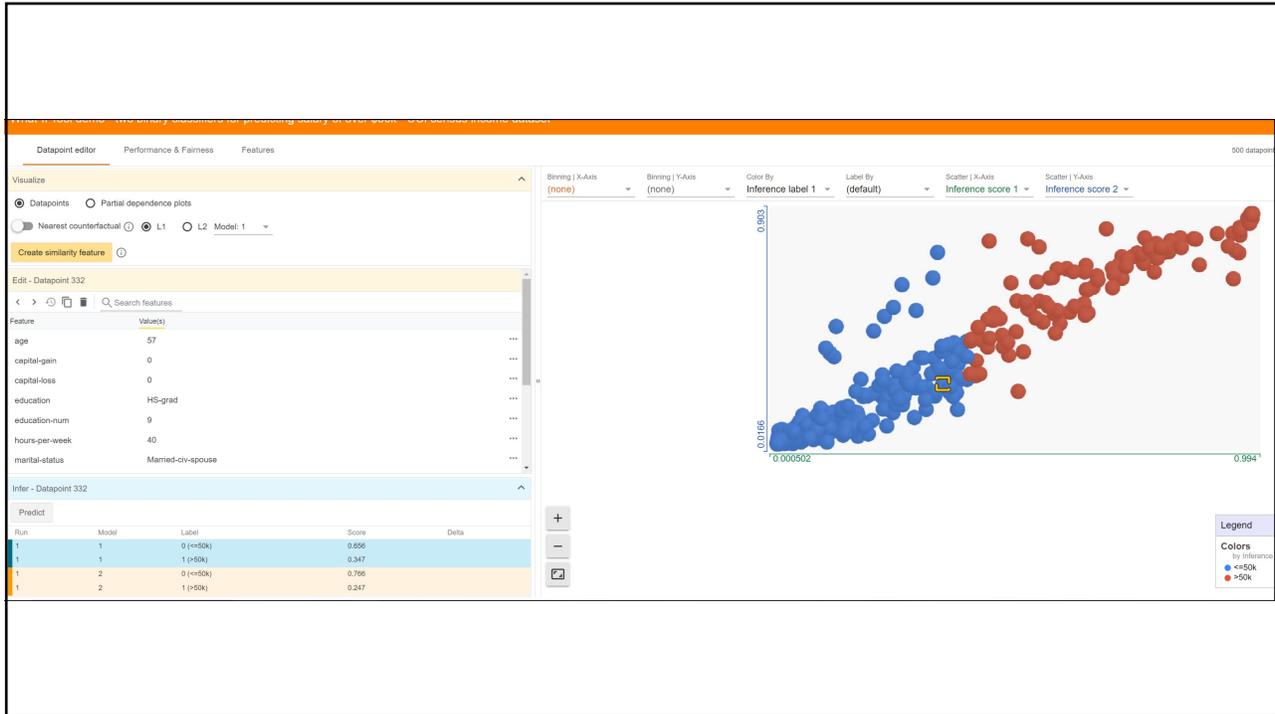


Wexler et al. (2019)

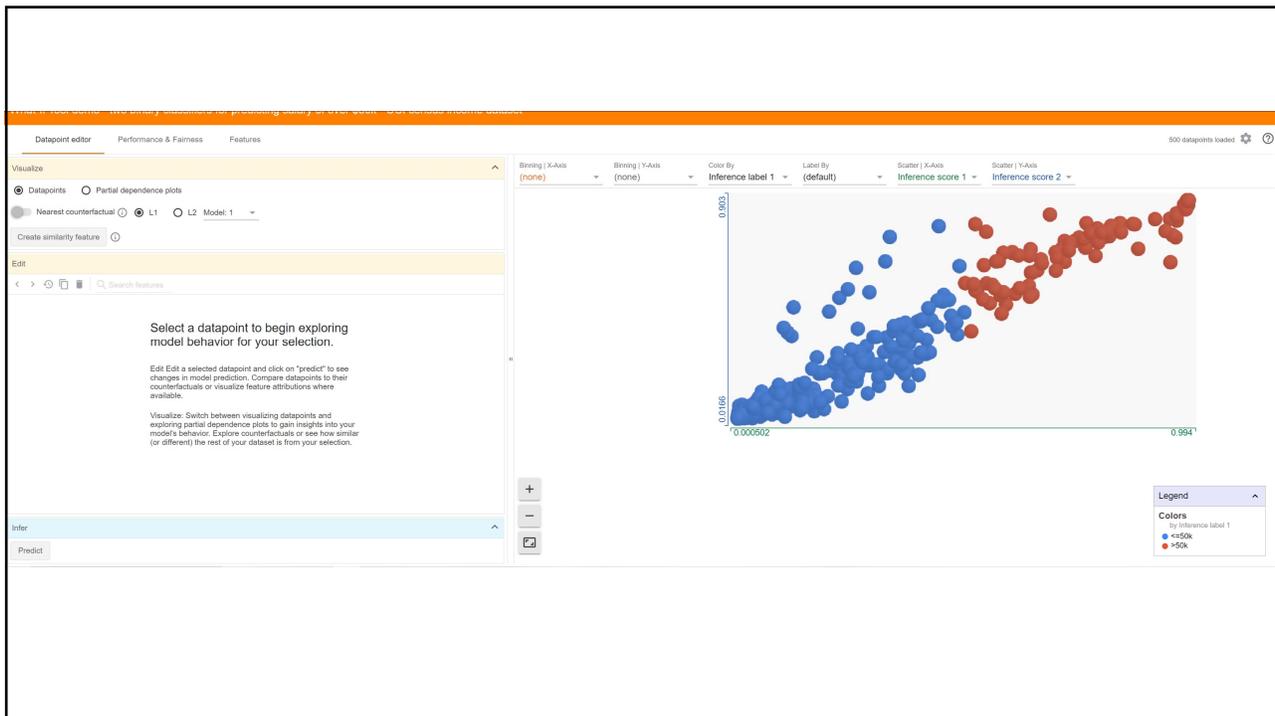
51



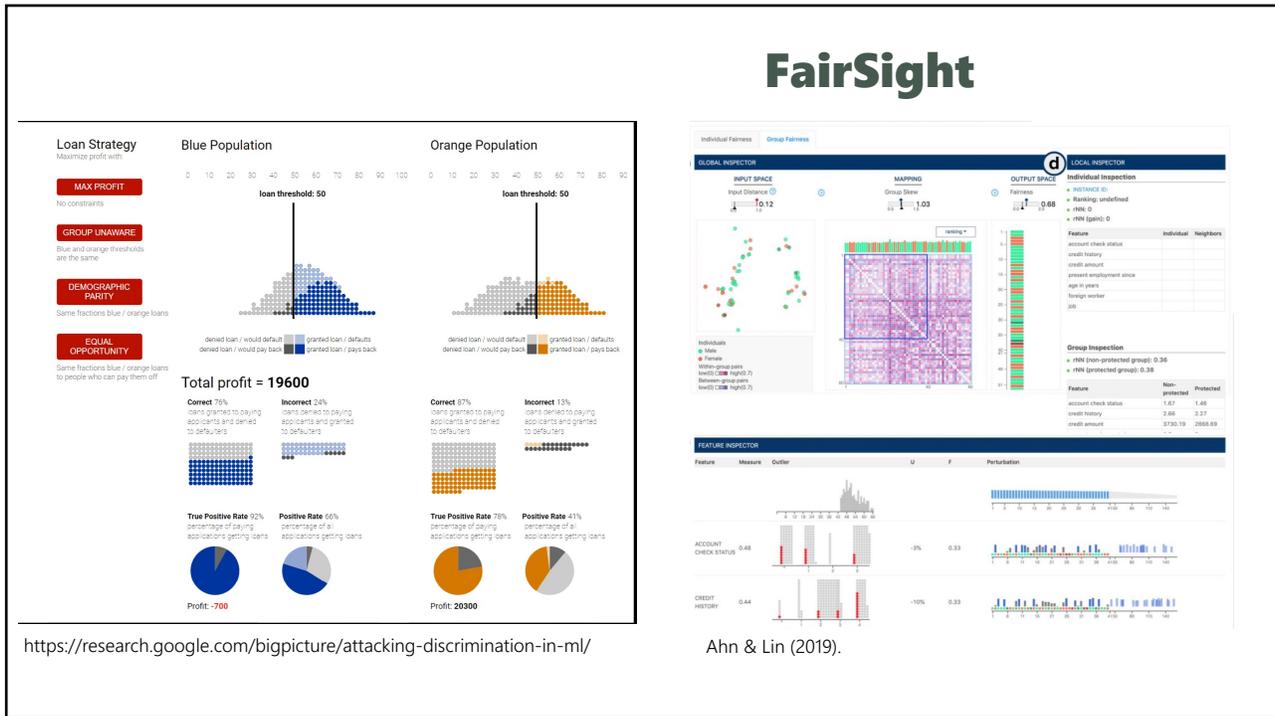
52



53



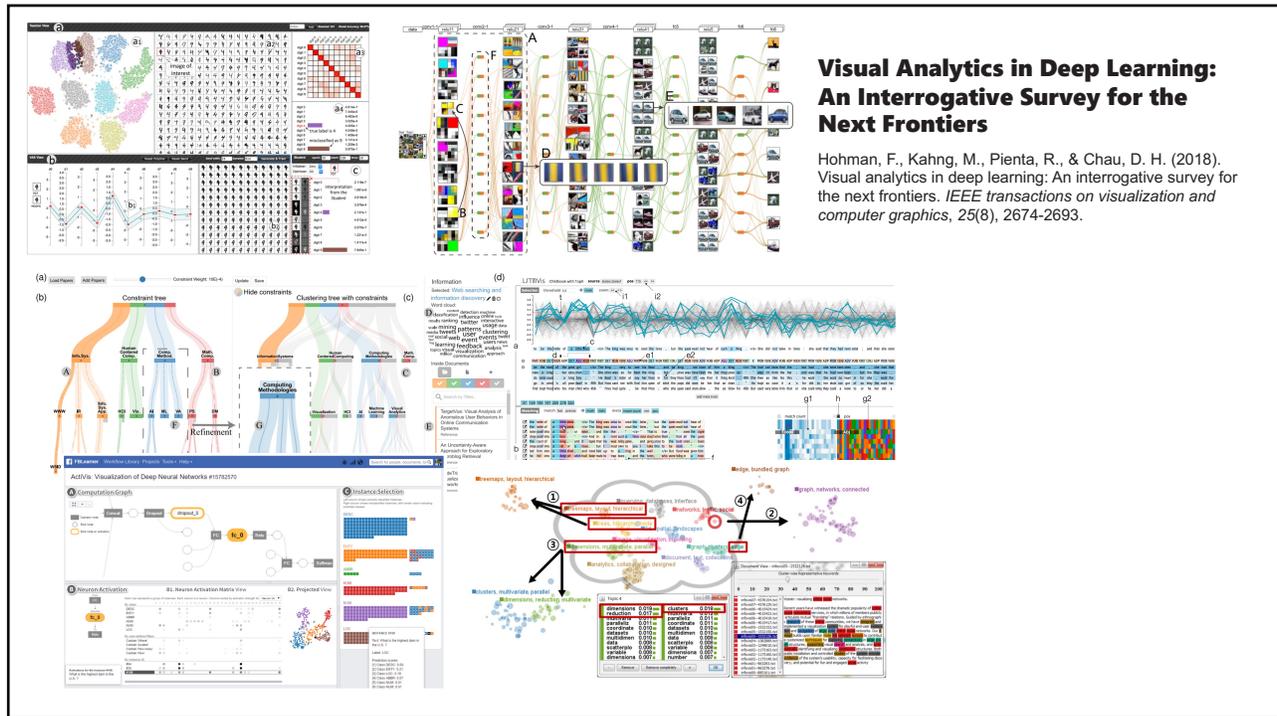
54



55

What is different/unique about visualizing ML data?

56



57

Some Guidelines for ML Visualizations

58

1. Visualizations should align with user expertise
2. Provide effective data representations for the task
3. Support understanding of model uncertainty
4. Exploit interactivity and promote rich interactions
5. Support expressive inputs

- Model Developers and Builders
- Model Users
- Domain Experts
- Non-Experts
- Learners/Students

	Predicted Yes	Predicted No	Total
Actual Yes	13.2% (66)	9.8% (49)	23.0% (115)
Actual No	5.2% (26)	71.8% (359)	77.0% (385)
Total	18.4% (92)	81.6% (408)	

Dudley & Kristensson (2018)

59

1. Visualizations should align with user expertise
2. Provide effective data representations for the task
3. Support understanding of model uncertainty
4. Exploit interactivity and promote rich interactions
5. Support expressive inputs

- Debugging and Improving Models
- Comparing and Selecting Models
- Interpretability and Explainability
- Teaching ML Concepts

"You mean to imply that I have nothin contrary, I can supply you with every dinner parties," warmly replied Chich spoke to prove his own rectitude and animated by the same desire.

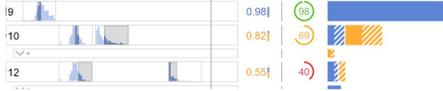
Kutuzov, shrugging his shoulders, rep smile: "I meant merely to say what I

Dudley & Kristensson (2018)

60

1. Visualizations should align with user expertise
2. Provide effective data representations for the task
3. Support understanding of model uncertainty
4. Exploit interactivity and promote rich interactions
5. Support expressive inputs

Uncertainty is an inevitable feature of data-driven models in most real-world applications.



Jessica Hullman
Northwestern University
Theories of inference for visual analysis
December 3, 2021

Research and development in computer science and statistics have produced increasingly sophisticated software interfaces for



Dudley & Kristensson (2018)

61

1. Visualizations should align with user expertise
2. Provide effective data representations for the task
3. Support understanding of model uncertainty
4. Exploit interactivity and promote rich interactions
5. Support expressive inputs

- Editing data points
- Evaluating Hypotheses
- Constructing Explanations

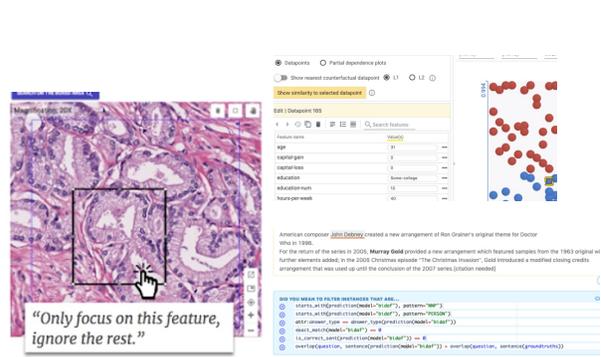


Dudley & Kristensson (2018)

62

1. Visualizations should align with user expertise
2. Provide effective data representations for the task
3. Support understanding of model uncertainty
4. Exploit interactivity and promote rich interactions
5. Support expressive inputs

- Direct Manipulation
- Query-by-demonstration



Dudley & Kristensson (2018)

63

Additional Resources

A visual introduction to machine learning

<http://www.r2d3.us/visual-intro-to-machine-learning-part-1/>

How to Use t-SNE Effectively

Although extremely useful for visualizing high-dimensional data, t-SNE plots can sometimes be mysterious or misleading. By exploring how it behaves in simple cases, we can learn to use it more effectively.

<https://distill.pub/2016/misread-tsne/>

The Building Blocks of Interpretability

Interpretability techniques are normally studied in isolation.

We explore the powerful interfaces that arise when you combine them — and the rich structure of this combinatorial space.

<https://distill.pub/2018/building-blocks/>

But what is a Neural Network?

An overview of what a neural network is, introduced in the context of recognizing hand-written digits.

<https://www.3blue1brown.com/topics/neural-networks>

Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers

Hohman, F., Kahng, M., Pienta, R., & Chau, D. H. (2018). Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE transactions on visualization and computer graphics*, 25(8), 2674-2693.

64