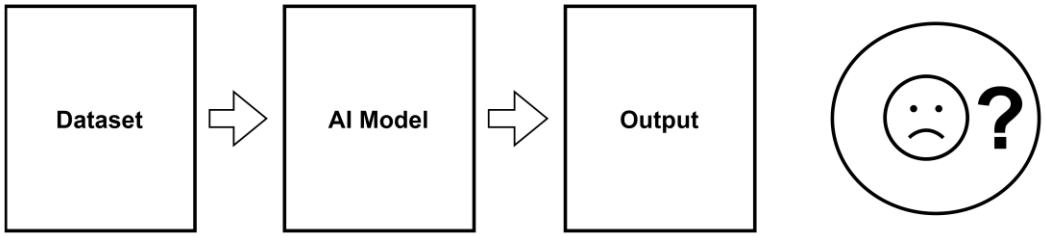
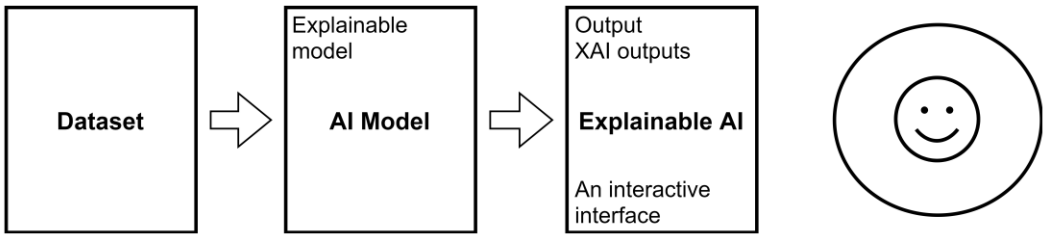
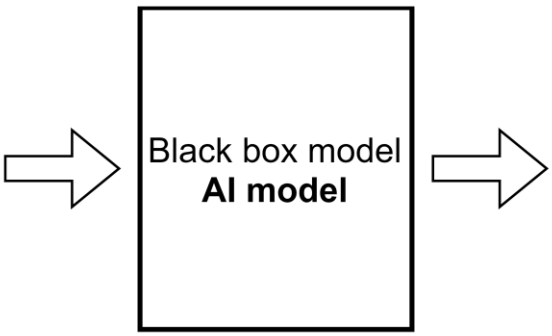
Chapter 01: Explaining Artificial Intelligence with Python

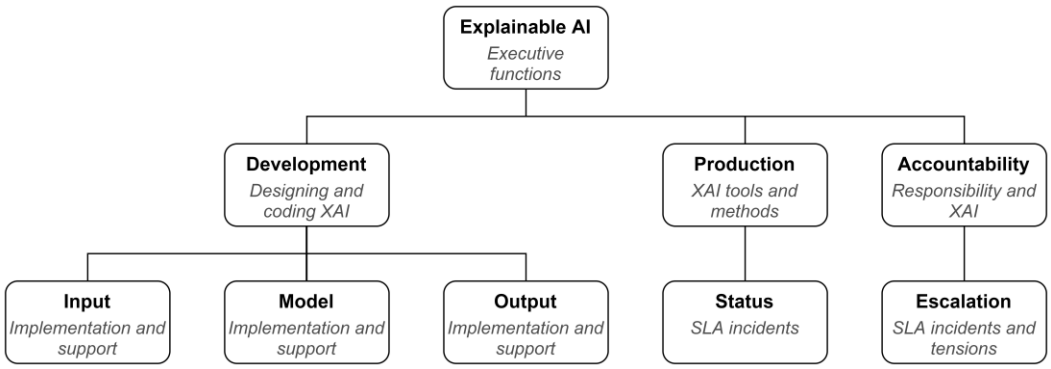
Artificial intelligence with little to no AI explanation

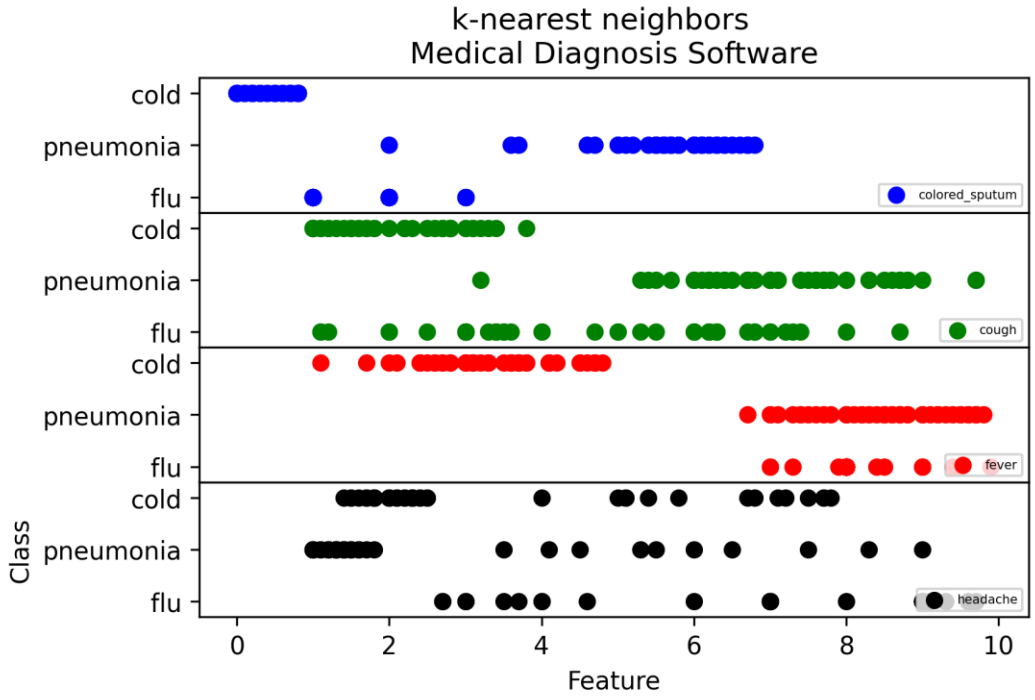


Artificial intelligence with AI explaining interface






Explainable AI
Executive functions
Development
Designing and coding XAI
Production
XAI tools and methods
Accountability
Responsibility and XAI
Input
Implementation and support
Model
Implementation and support
Output
Implementation and support
Status
SLA incidents
Escalation
SLA incidents and tensions


k-nearest neighbors
Medical Diagnosis Software
cold
pneumonia
flu
colored_sputum
cough
fever
headache
Class
0
2
4
6
8
10
Feature

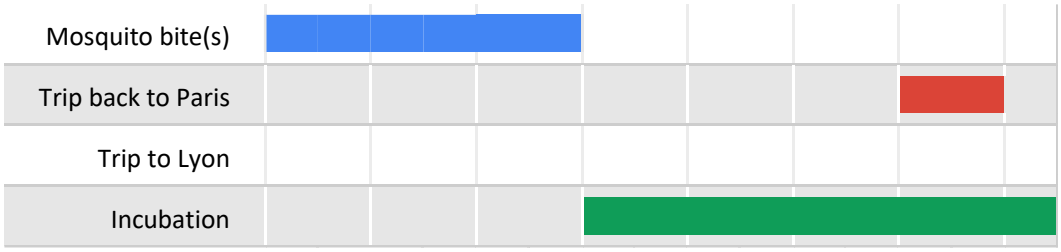
Evaluation Form

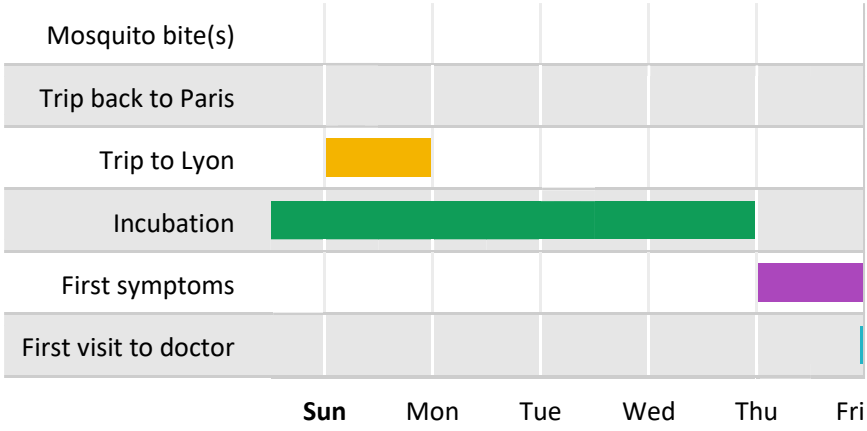
colored_sputum: 1

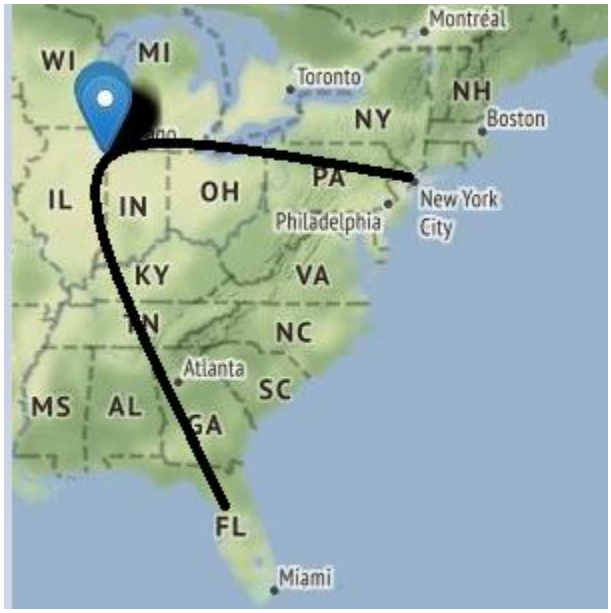
cough: 3

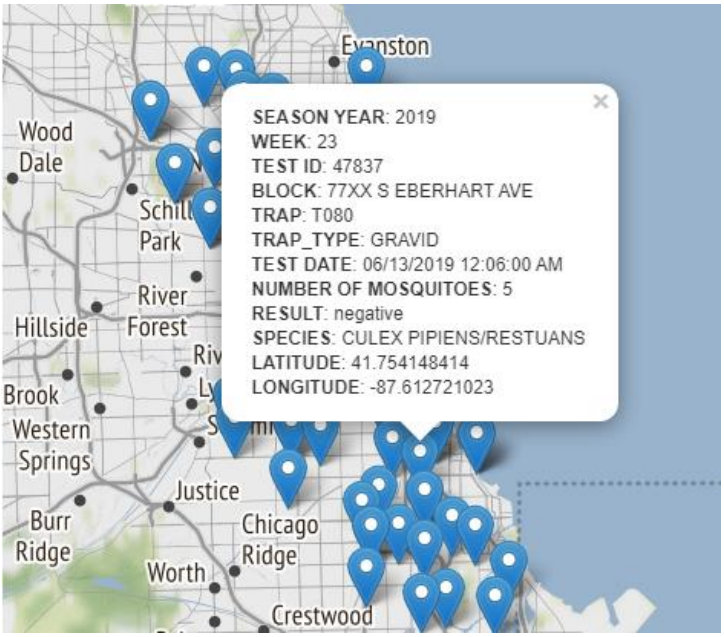
fever: 7

headache: 5

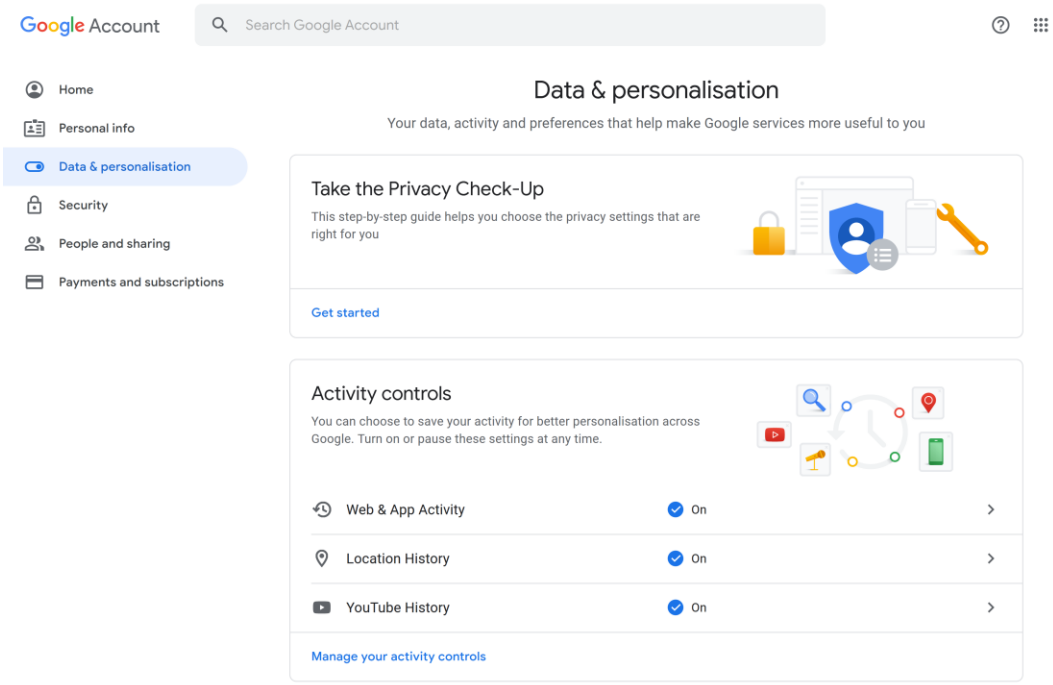


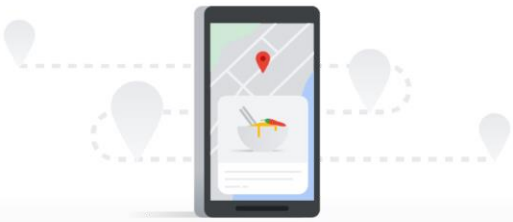



Montréal
WI
MI
Toronto
NH
NY
Boston
IL
IN
OH
PA
New York City
Philadelphia
KY
VA
NC
Atlanta
SC
MS
AL
GA
FL
Miami


Evanston
Wood Dale
Park
River Forest
Hillside
Brook
Western Springs
Justice
Burr Ridge
Chicago Ridge
Worth
Crestwood
SEASON YEAR: 2019
WEEK: 23
TEST ID: 47837
BLOCK: 77XX S EBERHART AVE
TRAP: T080
TRAP_TYPE: GRAVID
TEST DATE: 06/13/2019 12:06:00 AM
NUMBER OF MOSQUITOES: 5
RESULT: negative
SPECIES: CULEX PIPIENS/RESTUANS
LATITUDE: 41.754148414
LONGITUDE: -87.612721023

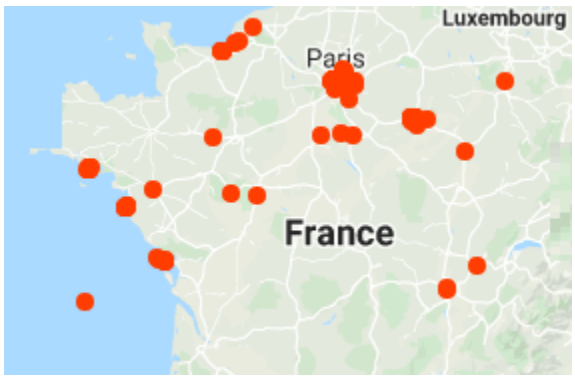
Bruss
LUXE
Paris
EURE
CÔTES-
D'ARMOR
AUBE
FRANCE
AVEYRON


Google Account
Search Google Account
Home
Personal info
Data & personalisation
Security
People and sharing
Payments and subscriptions
Data & personalisation
Your data, activity and preferences that help make Google services more useful to you
Take the Privacy Check-Up
This step-by-step guide helps you choose the privacy settings that are right for you
Get started
Activity controls
You can choose to save your activity for better personalisation across Google. Turn on or pause these settings at any time.
Web & App Activity
On
Location History
On
YouTube History
On
Manage your activity controls



Location History

Saves where you go with your devices, even when you aren't using a specific Google service, to give you personalised maps, recommendations based on places that you've visited and more. Learn more



1 Select data to include

Home App
Device, room, home and history information from the Home app. More info

Multiple formats

Keep
All notes and media attachments stored in Google Keep.

Multiple formats

Location History
Your Location History data collected while opted in to Location History.

Multiple formats

Location History

Your Location History data collected while opted in to Location History.

JSON and JSON formats

2 Choose file type, frequency and destination

Frequency:

Export once

1 export

Export every 2 months for 1 year

6 exports

File type & size

.zip

Zip files can be opened on almost any computer.

2 GB

Exports larger than this size will be split into multiple files.

Create export

Google

Your account, your data.

We've finished creating a copy of the Google data that you requested on February 3, 2020. You can download your files until February 10, 2020.

Your download will contain data from:

- Location History

Manage exports

Download your files

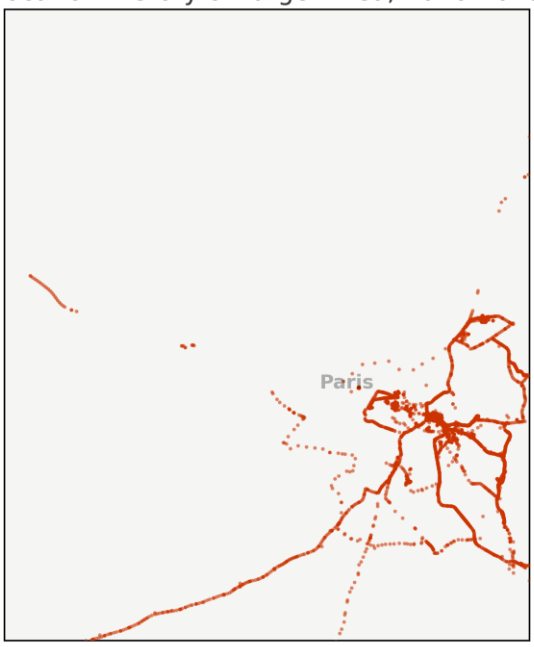
This message was sent to you because you recently used Google Takeout

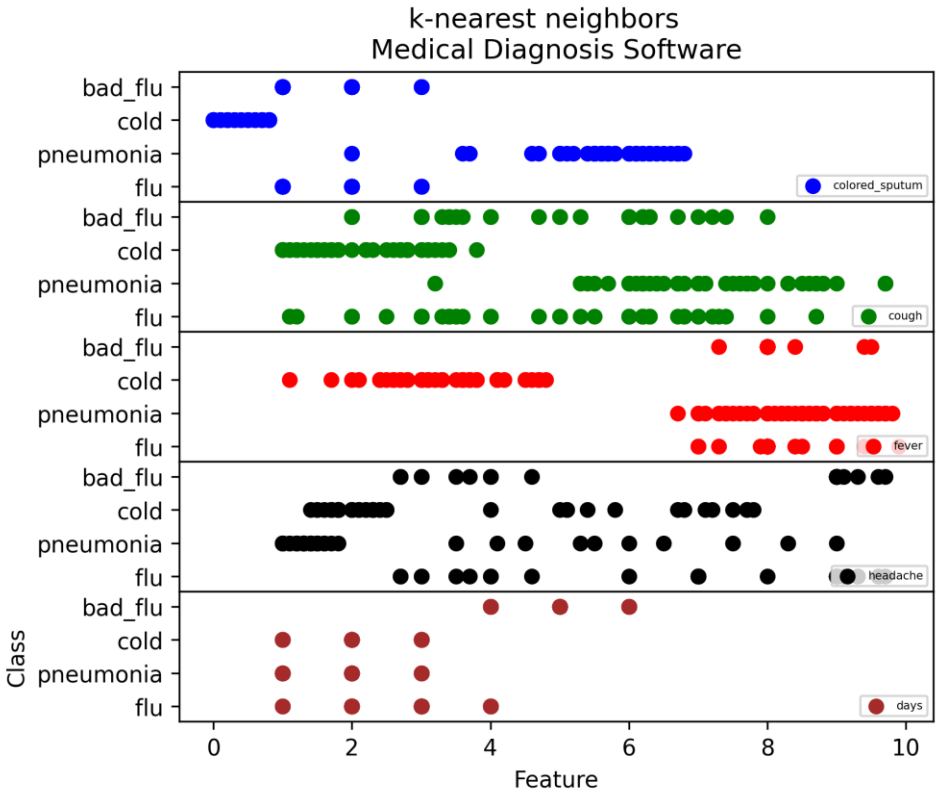
Privacy Policy | Terms of Service

Google

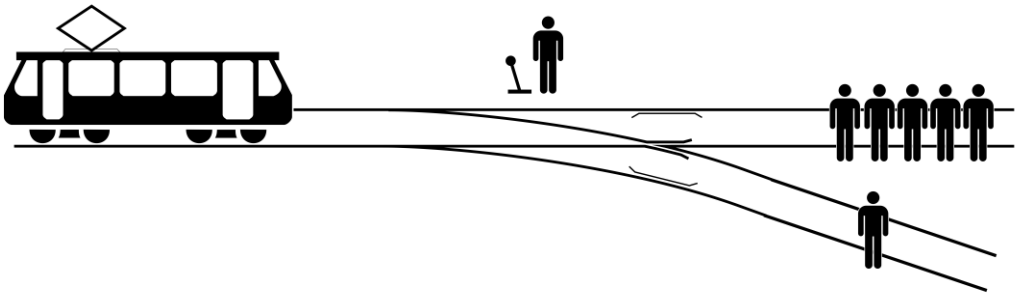
Export	Created on	Available until	Details
Location History less than 1 MB	February 3, 2020	February 10, 2020	Download

Location History of Target Area, 2016-2019

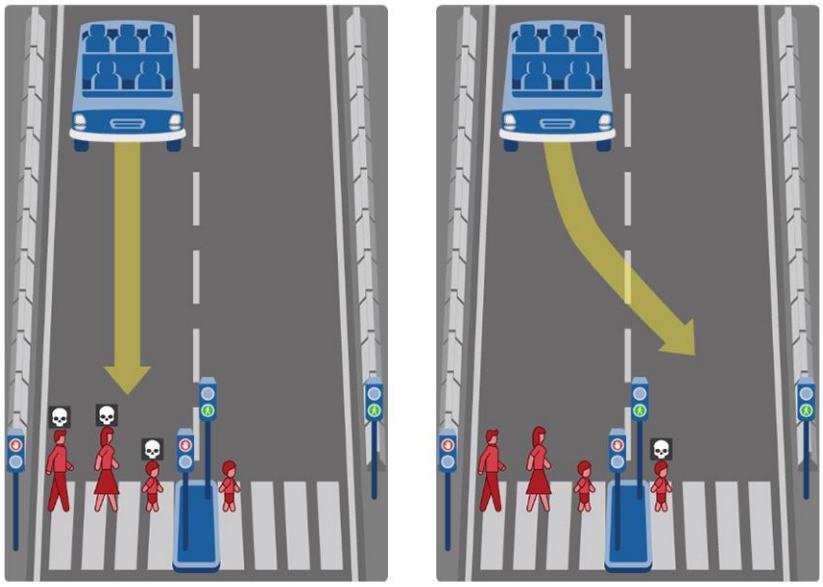


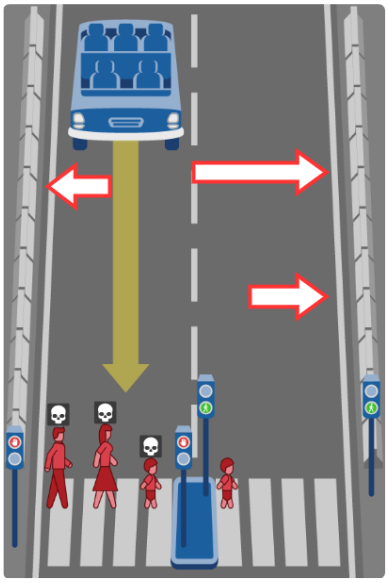

k-nearest neighbors
Medical Diagnosis Software
bad_flu
cold
pneumonia
flu
colored_sputum
cough
fever
headache
days
Class
0
2
4
6
8
10
Feature

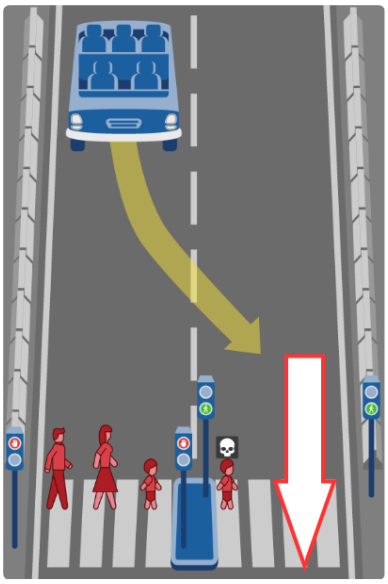
Chapter 02: White Box XAI for AI Bias and Ethics

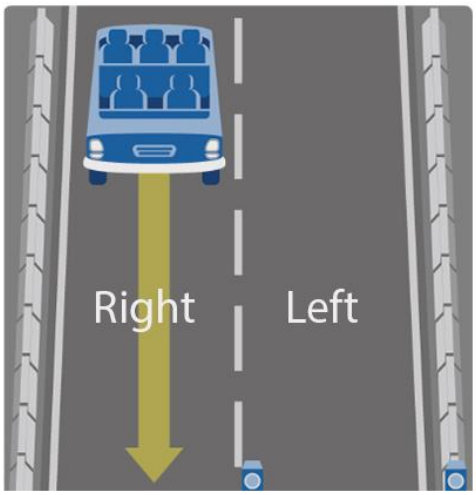


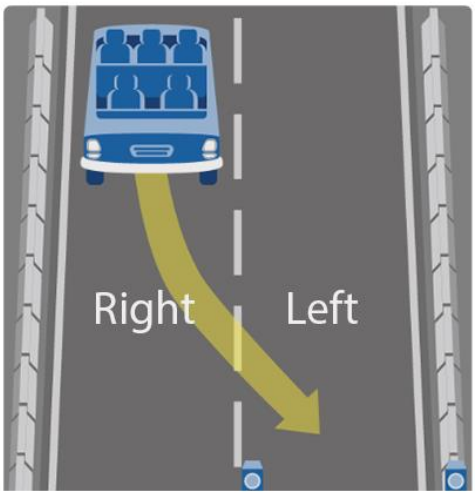
Uncertain speed of child pedestrian on other lane

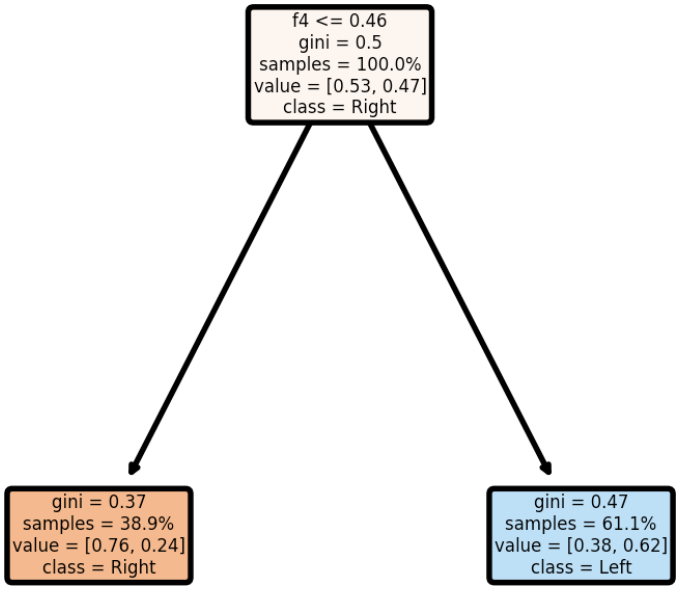


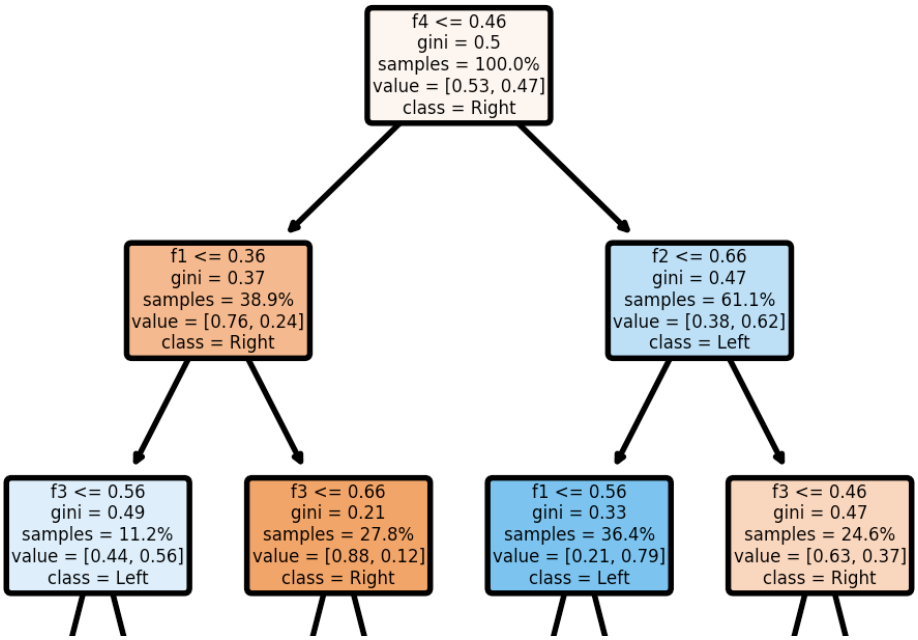


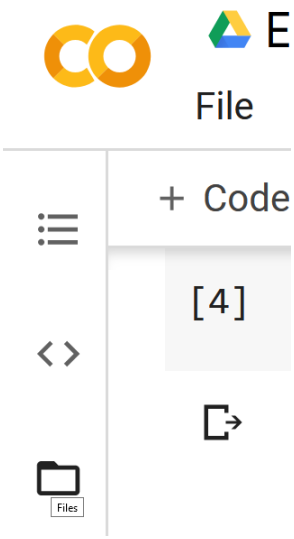


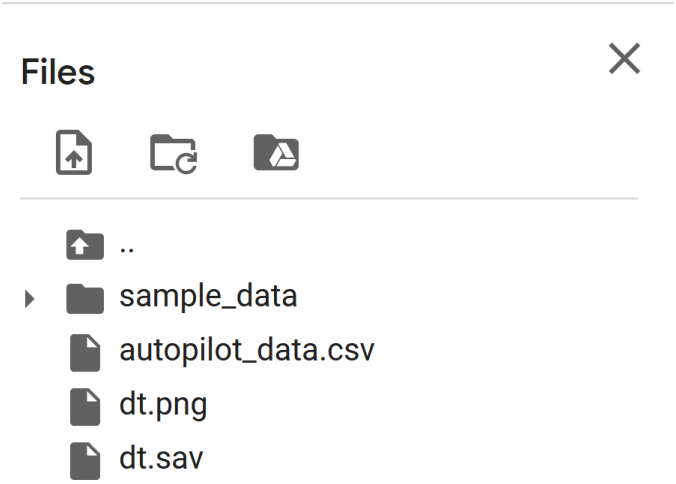

Right
Left

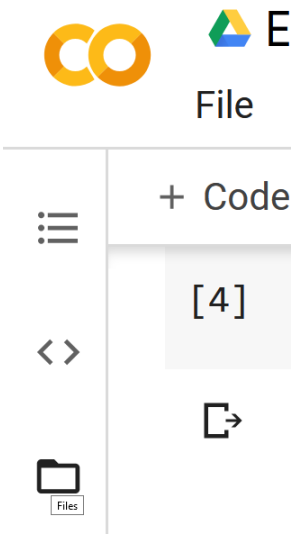

Right
Left


f4 <= 0.46
gini = 0.5
samples = 100.0%
value = [0.53, 0.47]
class = Right
gini = 0.37
samples = 38.9%
value = [0.76, 0.24]
class = Right
gini = 0.47
samples = 61.1%
value = [0.38, 0.62]
class = Left


f4 <= 0.46
gini = 0.5
samples = 100.0%
value = [0.53, 0.47]
class = Right
f1 <= 0.36
gini = 0.37
samples = 38.9%
value = [0.76, 0.24]
class = Right
f2 <= 0.66
gini = 0.47
samples = 61.1%
value = [0.38, 0.62]
class = Left
f3 <= 0.56
gini = 0.49
samples = 11.2%
value = [0.44, 0.56]
class = Left
f3 <= 0.66
gini = 0.21
samples = 27.8%
value = [0.88, 0.12]
class = Right
f1 <= 0.56
gini = 0.33
samples = 36.4%
value = [0.21, 0.79]
class = Left
f3 <= 0.46
gini = 0.47
samples = 24.6%
value = [0.63, 0.37]
class = Right


E
File
+ Code
[4]
Files


Files
..
sample_data
autopilot_data.csv
dt.png
dt.sav


E
File
+ Code
[4]
Files

Files

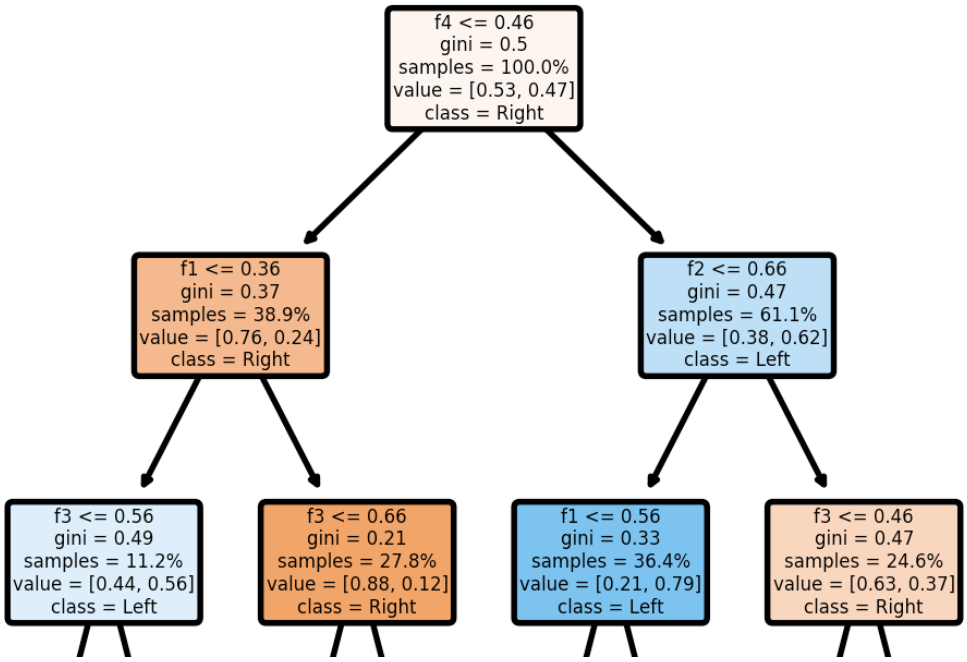
..

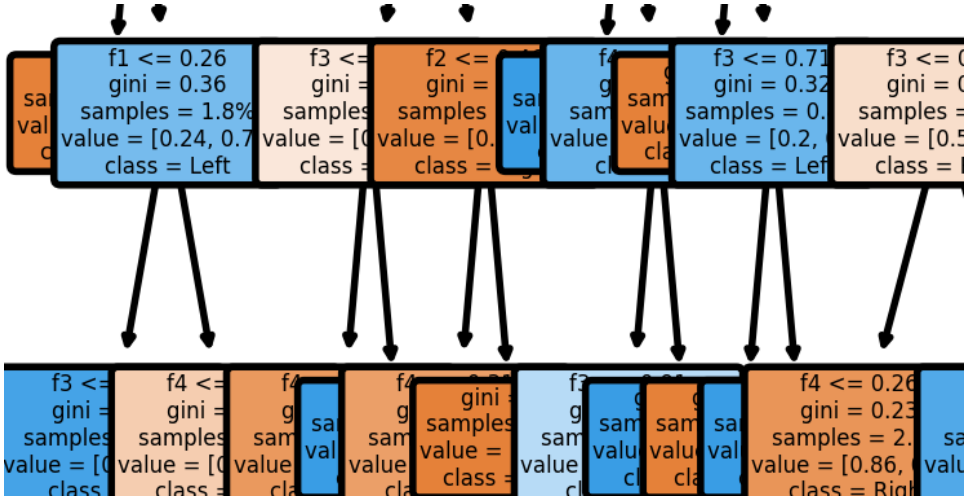
sample_data

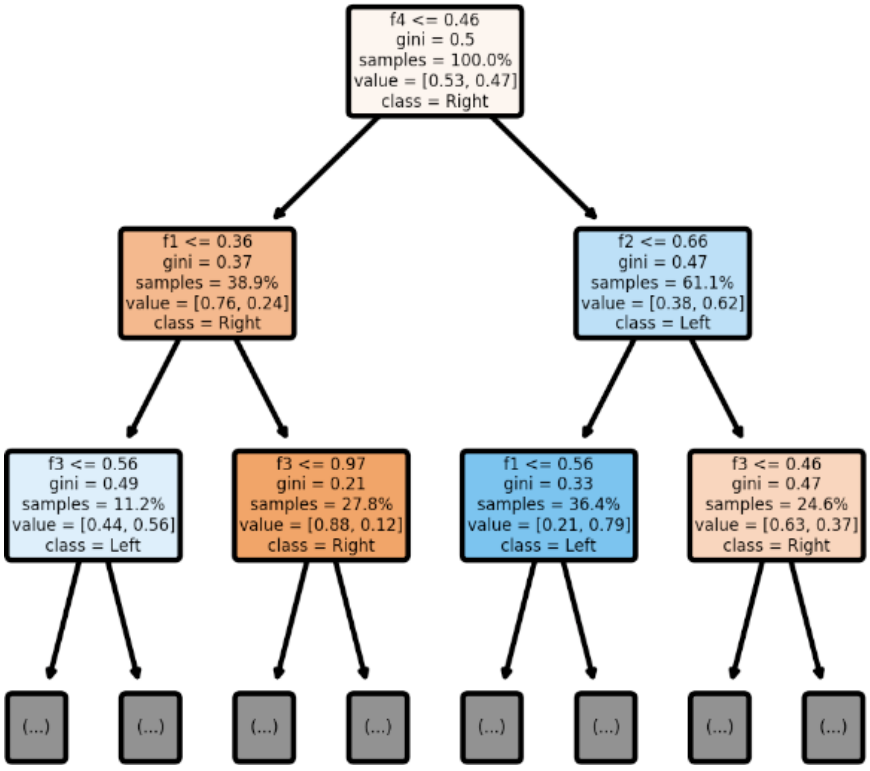
autopilot_data.csv

dt.png

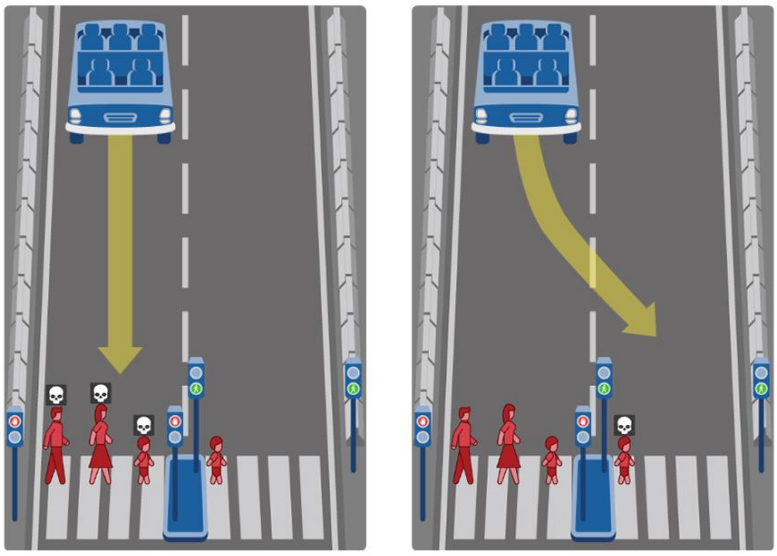
dt.sav




f1 <= 0.26
gini = 0.36
samples = 1.8%
class = Left
f3 <=
f2 <=
f3 <= 0.71
gini = 0.32
f3 <=
f3 <=
f4 <=
f4 <= 0.26
gini = 0.23


f4 <= 0.46
gini = 0.5
samples = 100.0%
value = [0.53, 0.47]
class = Right
f1 <= 0.36
gini = 0.37
samples = 38.9%
value = [0.76, 0.24]
class = Right
f2 <= 0.66
gini = 0.47
samples = 61.1%
value = [0.38, 0.62]
class = Left
f3 <= 0.56
gini = 0.49
samples = 11.2%
value = [0.44, 0.56]
class = Left
f3 <= 0.97
gini = 0.21
samples = 27.8%
value = [0.88, 0.12]
class = Right
f1 <= 0.56
gini = 0.33
samples = 36.4%
value = [0.21, 0.79]
class = Left
f3 <= 0.46
gini = 0.47
samples = 24.6%
value = [0.63, 0.37]
class = Right
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)

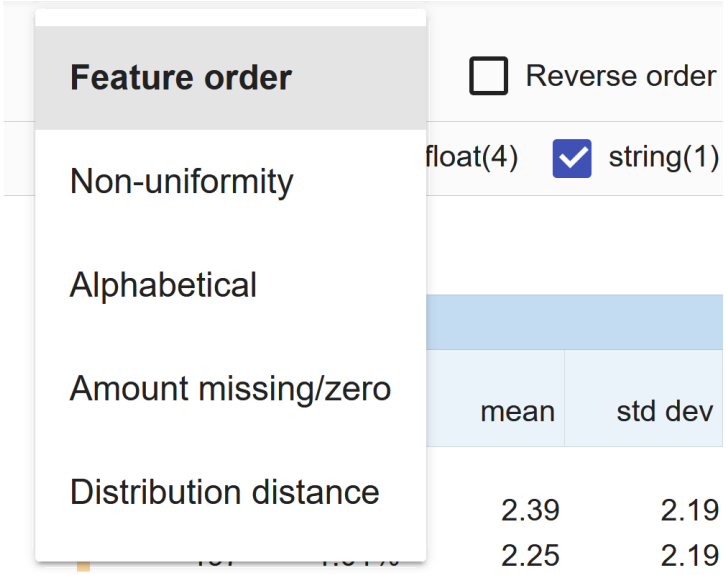
Uncertain speed of child pedestrian on other lane



Chapter 03: Explaining Machine Learning with Facets

Numeric Features (7)

	count	missing	mean	std dev	zeros
colored_sputum					
	184	0%	2.39	2.19	6.52%
	197	1.01%	2.25	2.19	8.12%
cough					
	184	0%	4.63	2.4	0%
	199	0%	4.42	2.42	0%
fever					
	184	0%	6.93	2.3	0%
	199	0%	6.61	2.44	0%



Sort by

Non-uniformity

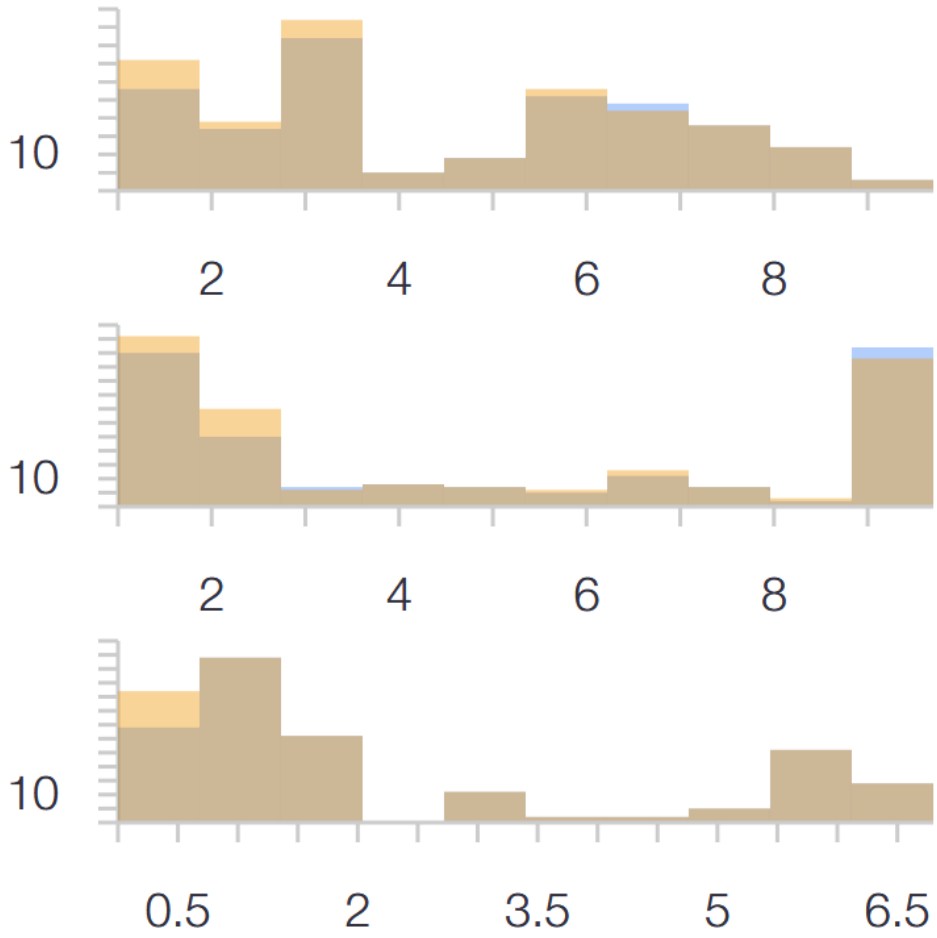


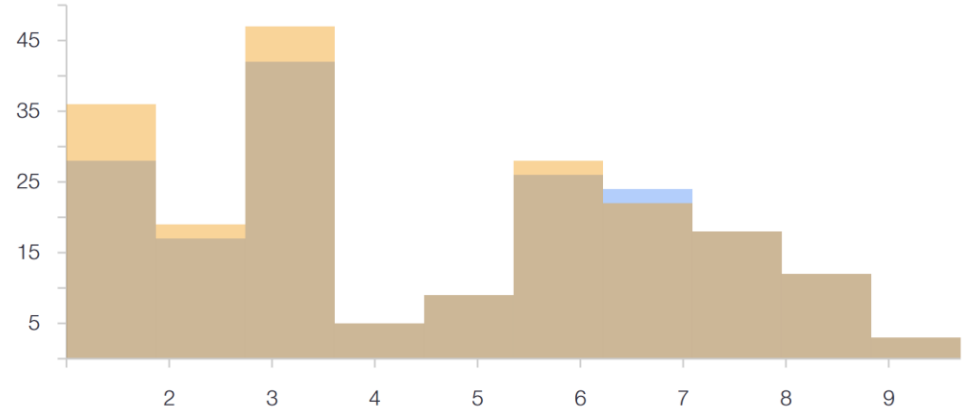
Chart to show

Standard

☐ log ☑ expand ☐ percentages

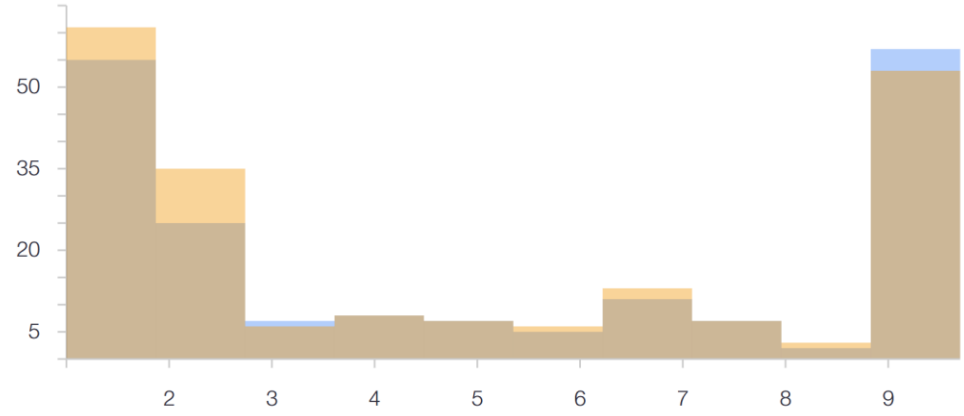
Numeric Features (7)

	count	missing	mean	std dev	zeros	min	median	max
cough								
	184	0%	4.63	2.4	0%	1	4.35	9.7
	199	0%	4.42	2.42	0%	1	3.5	9.7

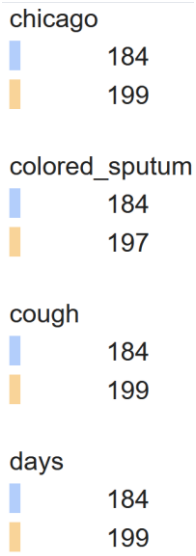


Numeric Features (7)

	count	missing	mean	std dev	zeros	min	median	max
headache								
	184	0%	4.97	3.27	0%	1	4	9.7
	199	0%	4.7	3.2	0%	1	3.5	9.7



	count	missing	mean	std dev	zeros	min	median	max
cough								
	184	0%	4.63	2.4	0%	1	4.35	9.7
	199	0%	4.42	2.42	0%	1	3.5	9.7
headache								
	184	0%	4.97	3.27	0%	1	4	9.7
	199	0%	4.7	3.2	0%	1	3.5	9.7

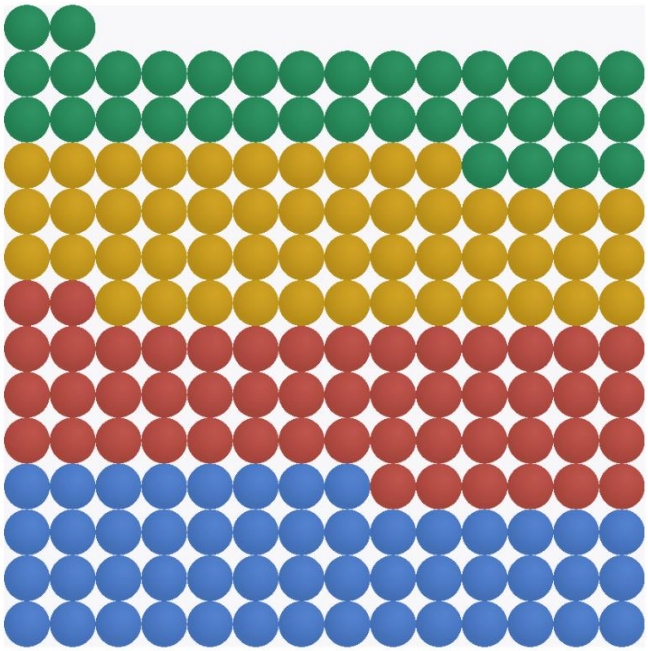


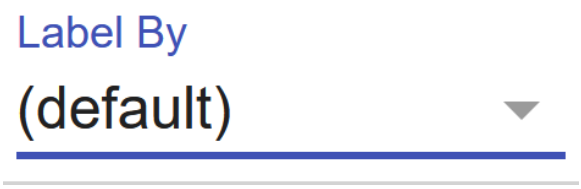
Numeric Features (7)

	count	missing	mean	std dev	zeros
france					
	184	0%	0	0	**100%**
	199	0%	0	0	**100%**
colored_sputum					
	184	0%	2.39	2.19	6.52%
	197	1.01%	2.25	2.19	8.12%

Numeric Features (7)					
	count	missing	mean	std dev	zeros
fever					
	184	0%	6.93	2.3	0%
	199	0%	6.61	2.44	0%
headache					
	184	0%	4.97	3.27	0%
	199	0%	4.7	3.2	0%
colored_sputum					
	184	0%	2.39	2.19	6.52%
	197	1.01%	2.25	2.19	8.12%

Binning | X-Axis
(none)
Binning | Y-Axis
(none)
Color By
class
Label By
(default)
Scatter | X-Axis
(default)
Scatter | Y-Axis
(default)

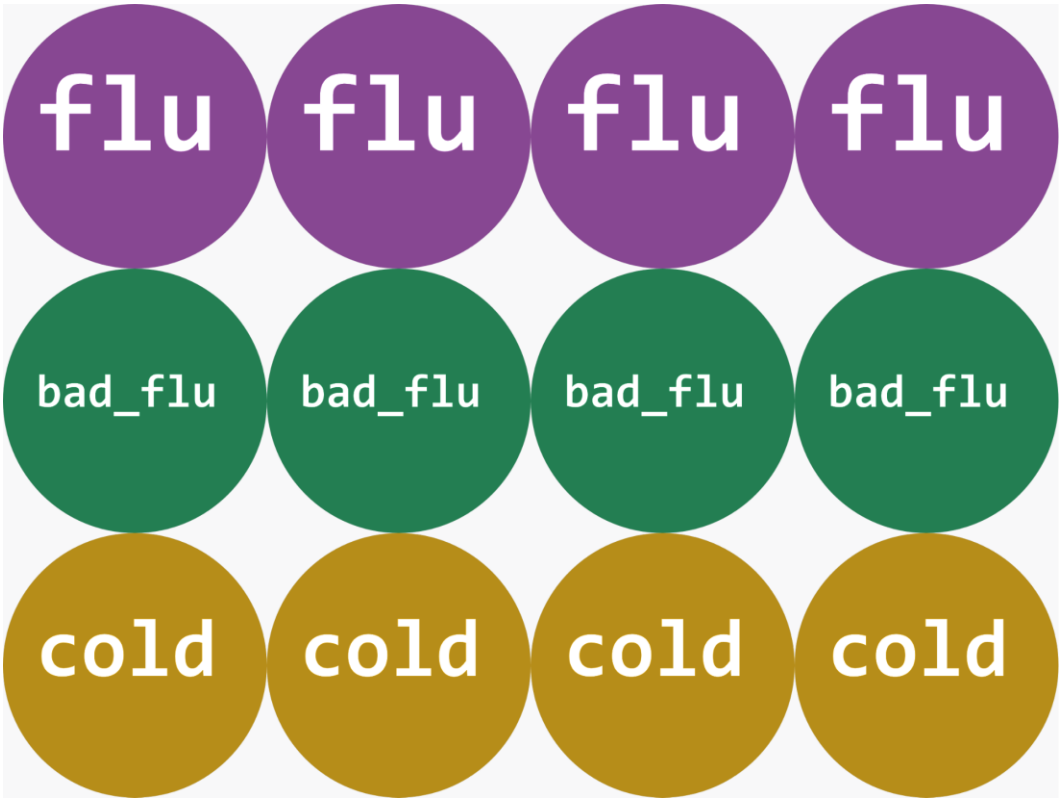



Label By
(default)

(default)
chicago
class
colored_sputum
cough
days
fever
france
headache

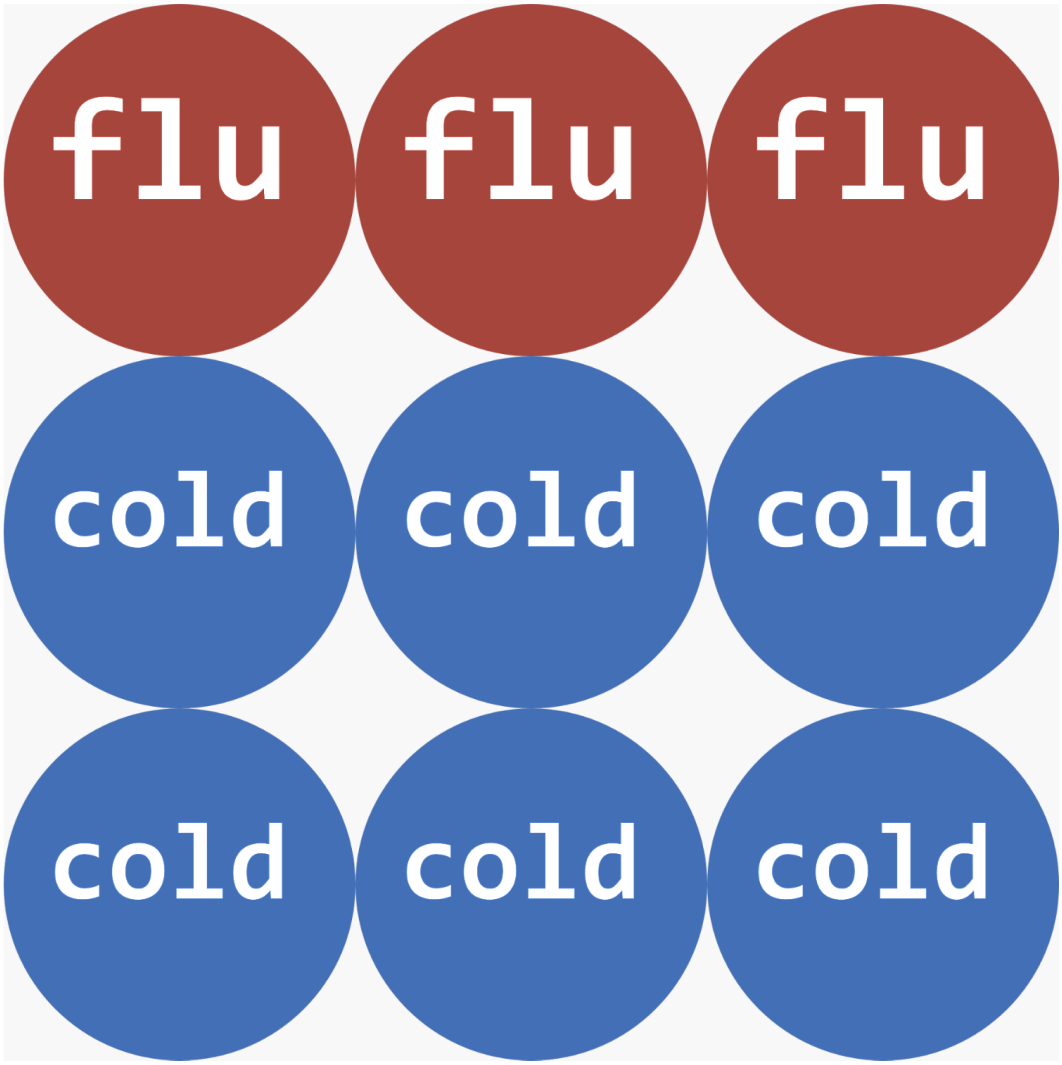
Label By

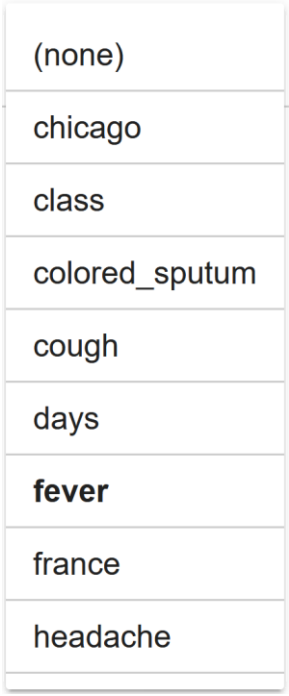
class



Color By

(none)


flu
flu
flu
cold
cold
cold
cold
cold
cold

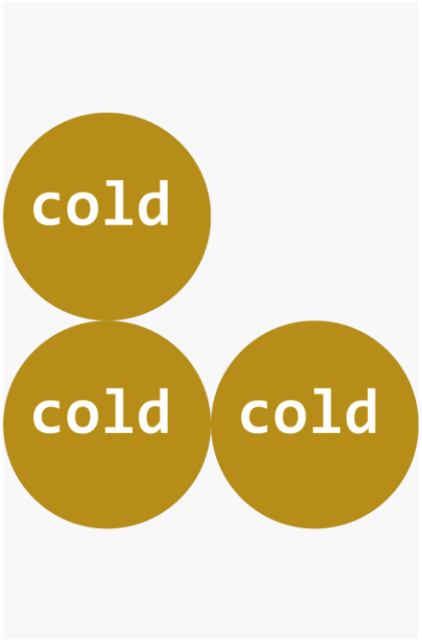


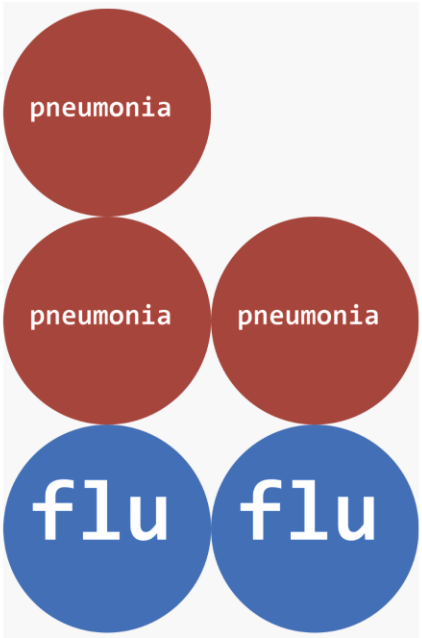
Binning | X-Axis

fever

Binning | Y-Axis

days





Scatter | X-Axis

(default)

Scatter | Y-Axis

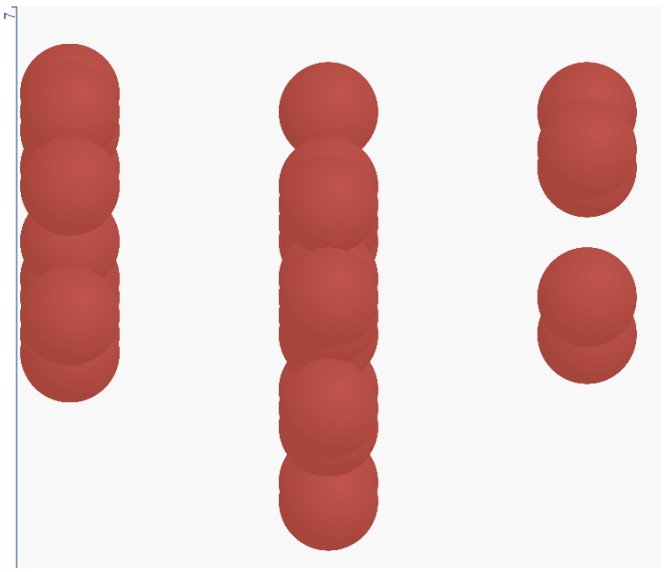
(default)

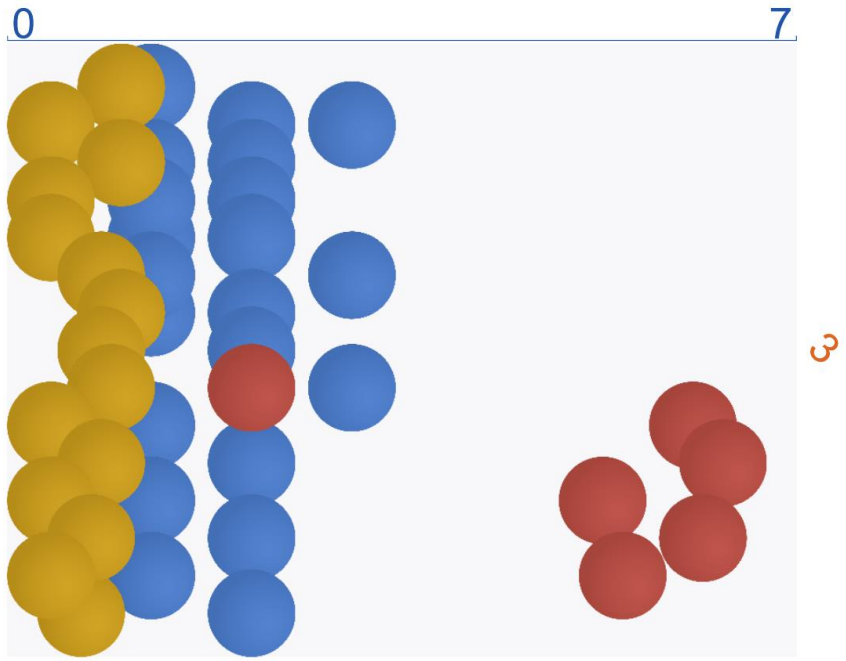
Scatter | X-Axis

days

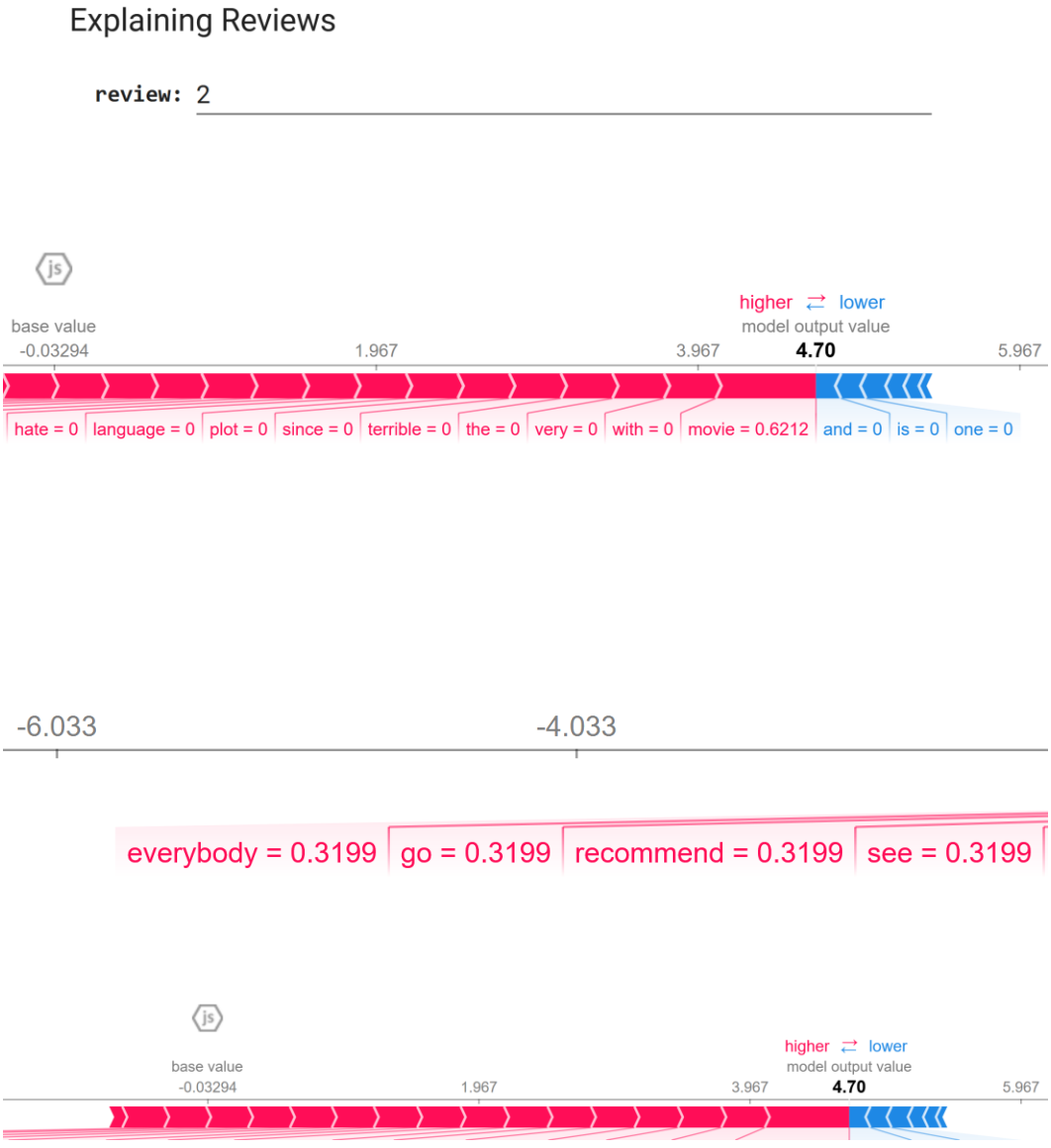
Scatter | Y-Axis

colored_sputu


Z


0
7
3

Chapter 04: Microsoft Azure Machine Learning Model Interpretability with SHAP



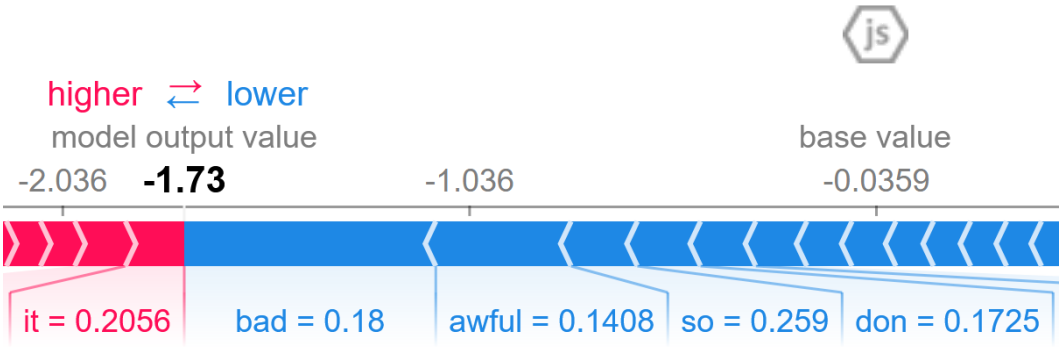
0 r2 True "Twelve Monkeys" is odd and disturbing

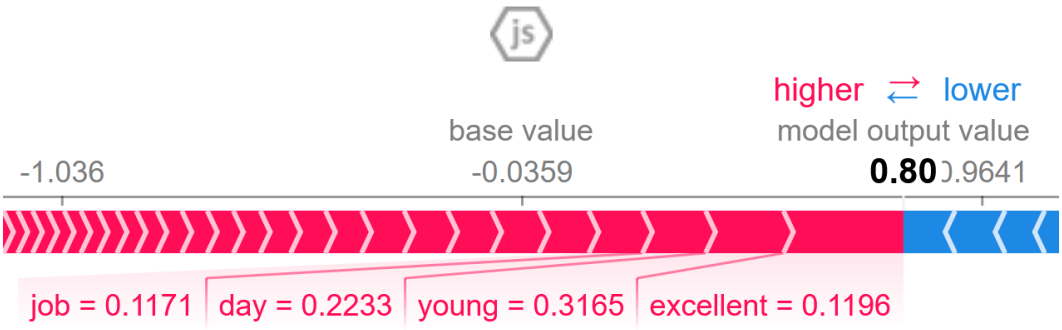
2 r1 True I finally saw this film tonight after

5 r5 False A chemical spill is turning people in

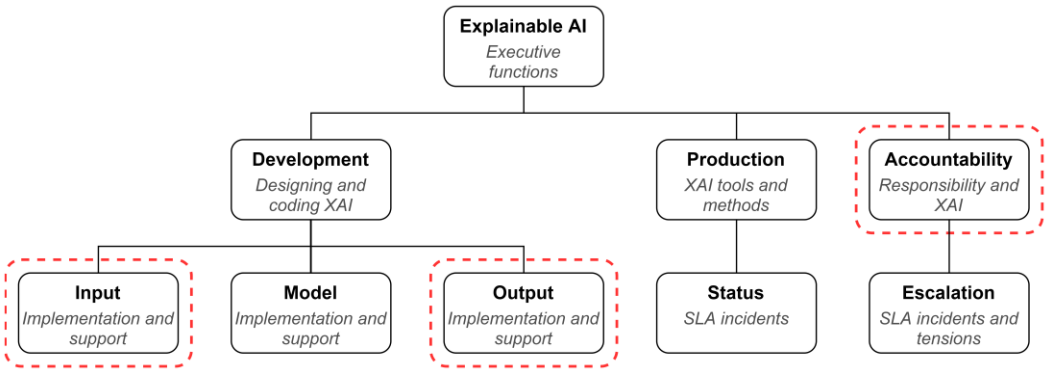
9 r4 True I first saw the trailer for Frailty on

59 r3 True I am currently on vacation in Israel


higher ⇄ lower
model output value
-2.036
-1.73
-1.036
base value
-0.0359
it = 0.2056
bad = 0.18
awful = 0.1408
so = 0.259
don = 0.1725


higher ⇄ lower
model output value
base value
-0.0359
-1.036
0.80
job = 0.1171
day = 0.2233
young = 0.3165
excellent = 0.1196

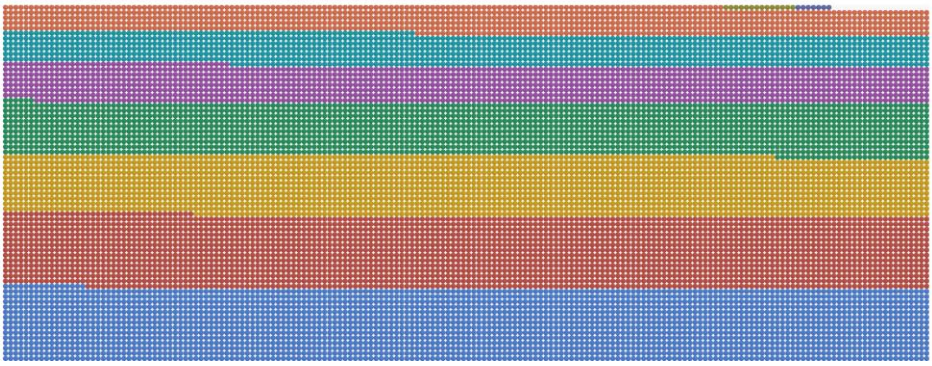
Chapter 05: Building an Explainable AI Solution from Scratch



	Age	Workclass	fnlwgt	Education	Education-Num	Marital-Status	Occupation
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners

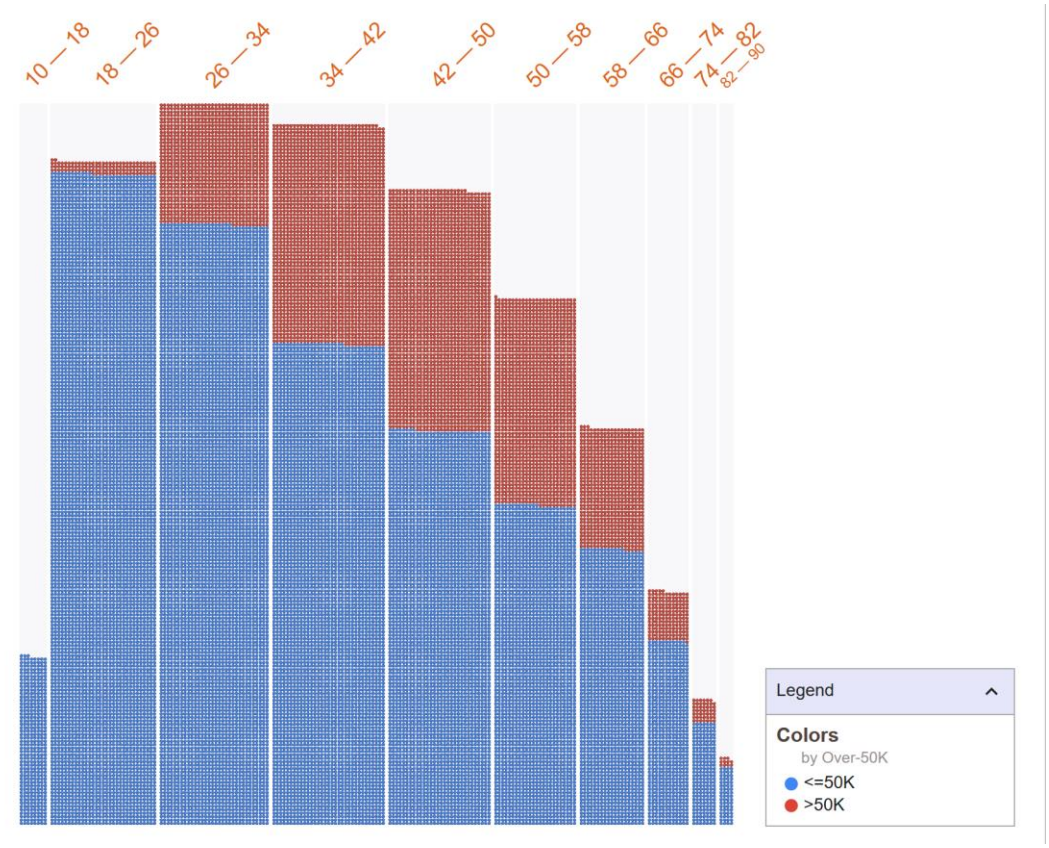
Relationship	Race	Sex	Capital-Gain	Capital-Loss	Hours-per-week	Country	Over-50K
Not-in-family	White	Male	2174	0	40	United-States	<=50K
Husband	White	Male	0	0	13	United-States	<=50K
Not-in-family	White	Male	0	0	40	United-States	<=50K
Husband	Black	Male	0	0	40	United-States	<=50K

Binning | X-Axis (none) | Binning | Y-Axis (none) | Color By Workclass | Label By Education | Scatter | X-Axis (default) | Scatter | Y-Axis (default)



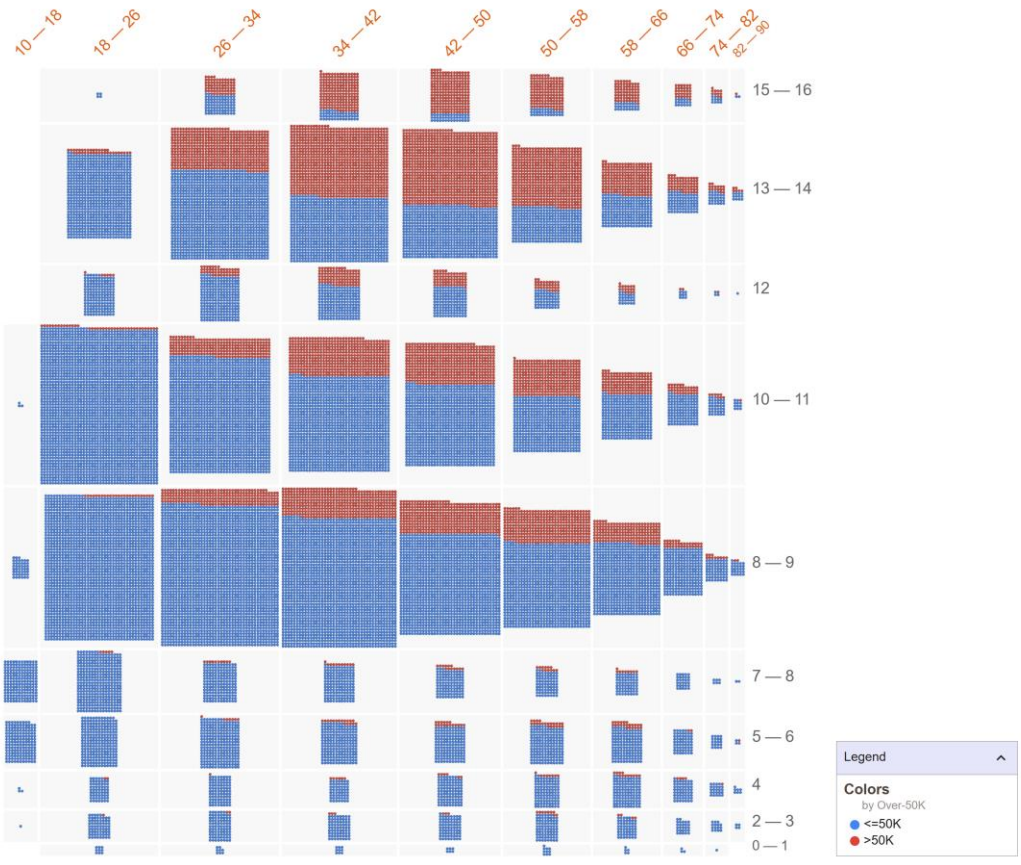
Display the Dive visualization for the training data.

Binning | X-Axis Age | Count 10 | Binning | Y-Axis (none) | Color By Over-50K

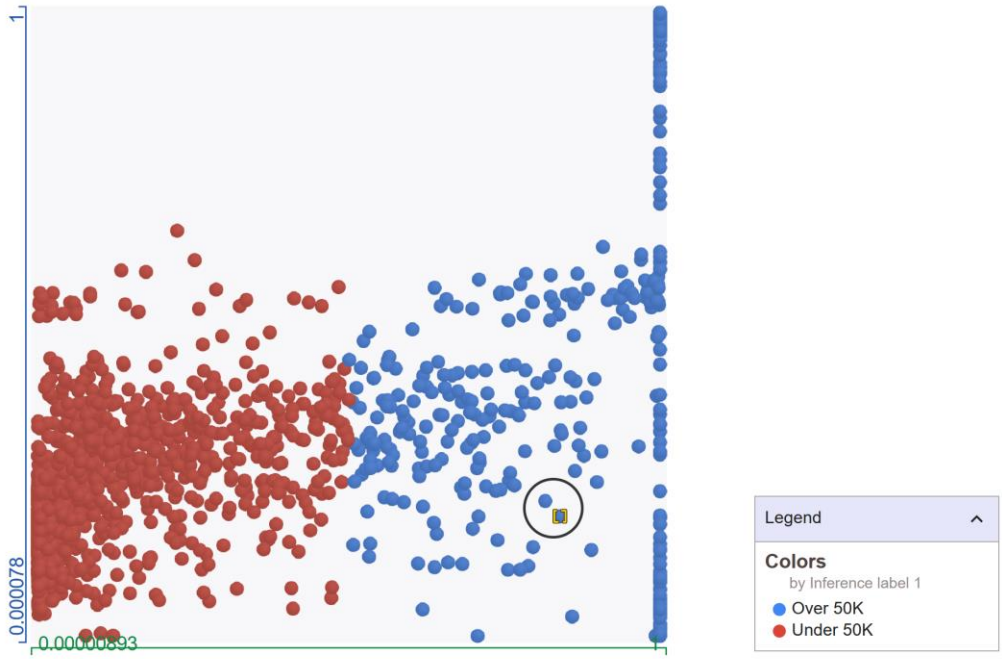


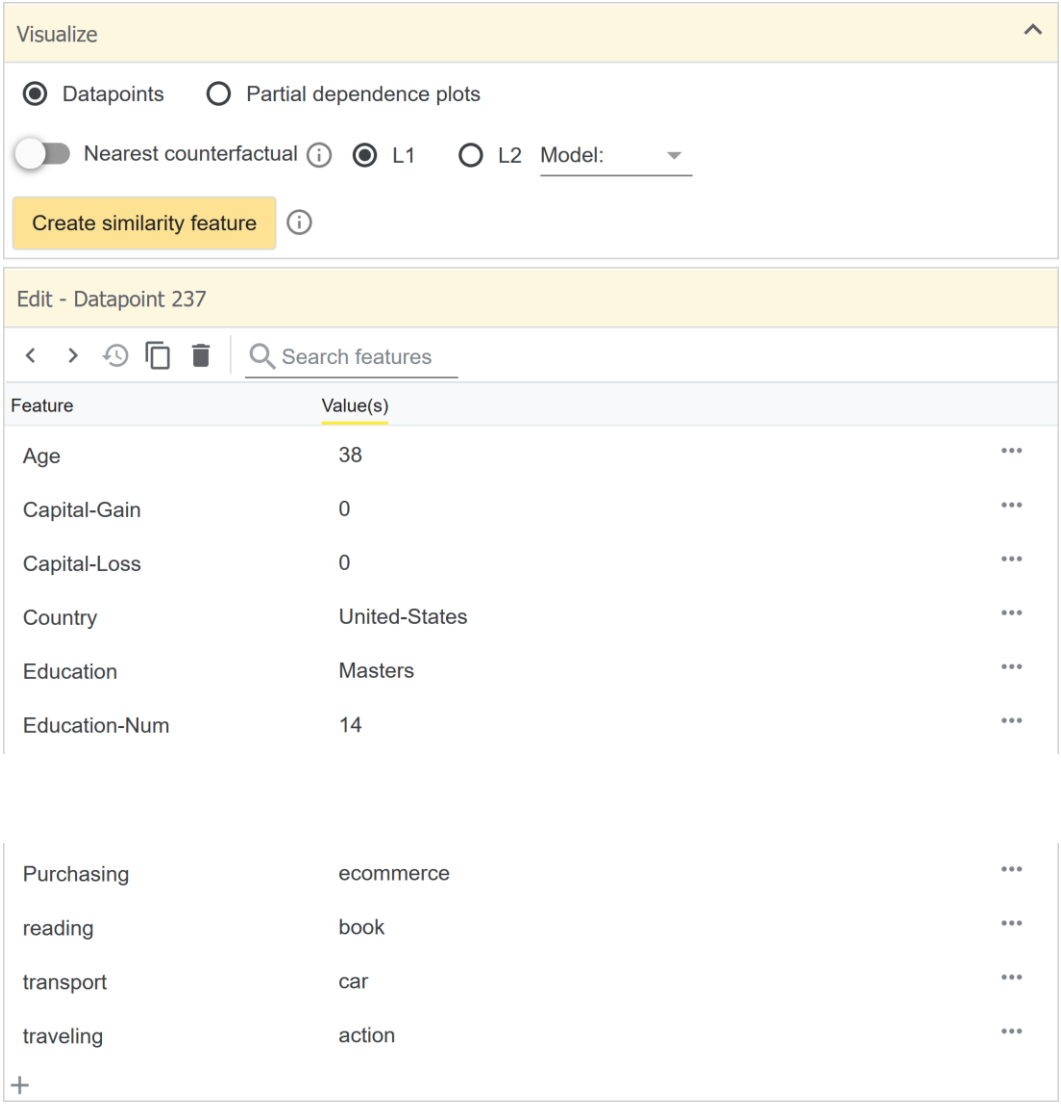
Display the Dive visualization for the training data.

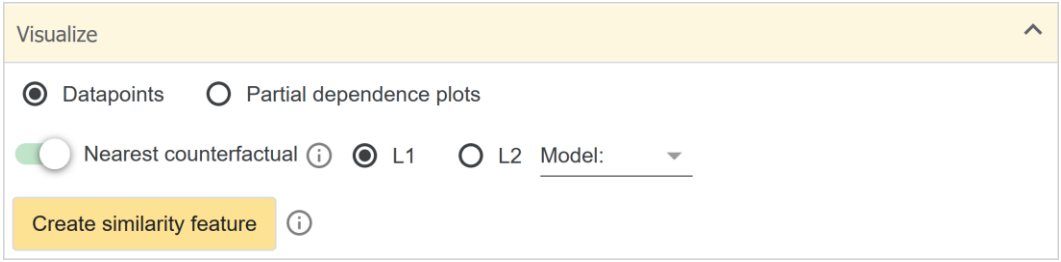
Binning | X-Axis	Count	Binning | Y-Axis	Count	Color By
Age	10	Education-Num	10	Over-50K

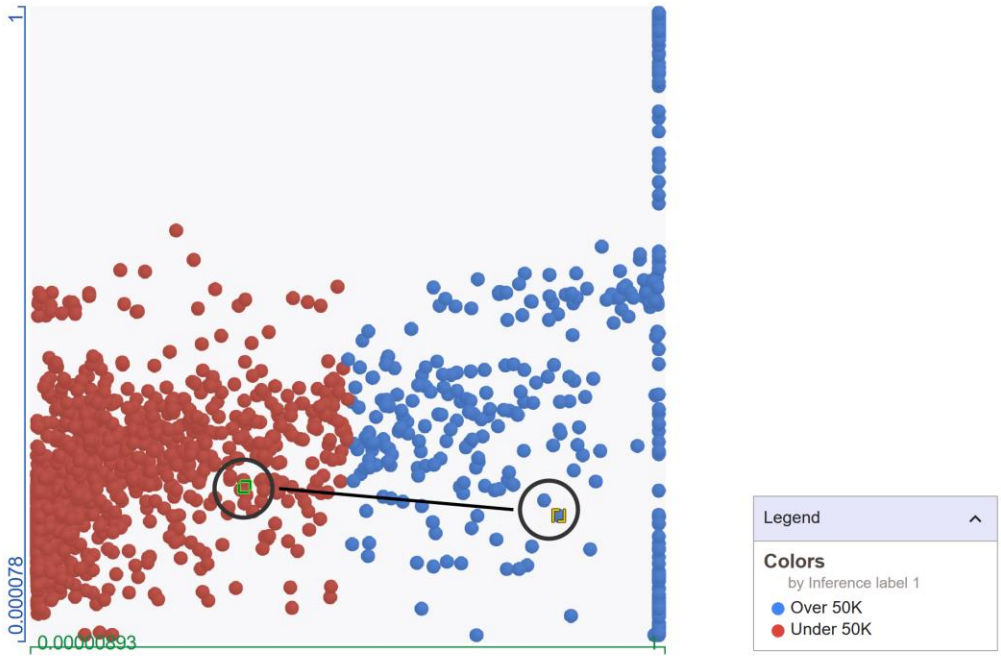

10 — 18
18 — 26
26 — 34
34 — 42
42 — 50
50 — 58
58 — 66
66 — 74
74 — 82
82 — 90
15 — 16
13 — 14
12
10 — 11
8 — 9
7 — 8
5 — 6
4
2 — 3
0 — 1
Legend
Colors
by Over-50K
<=50K
>50K

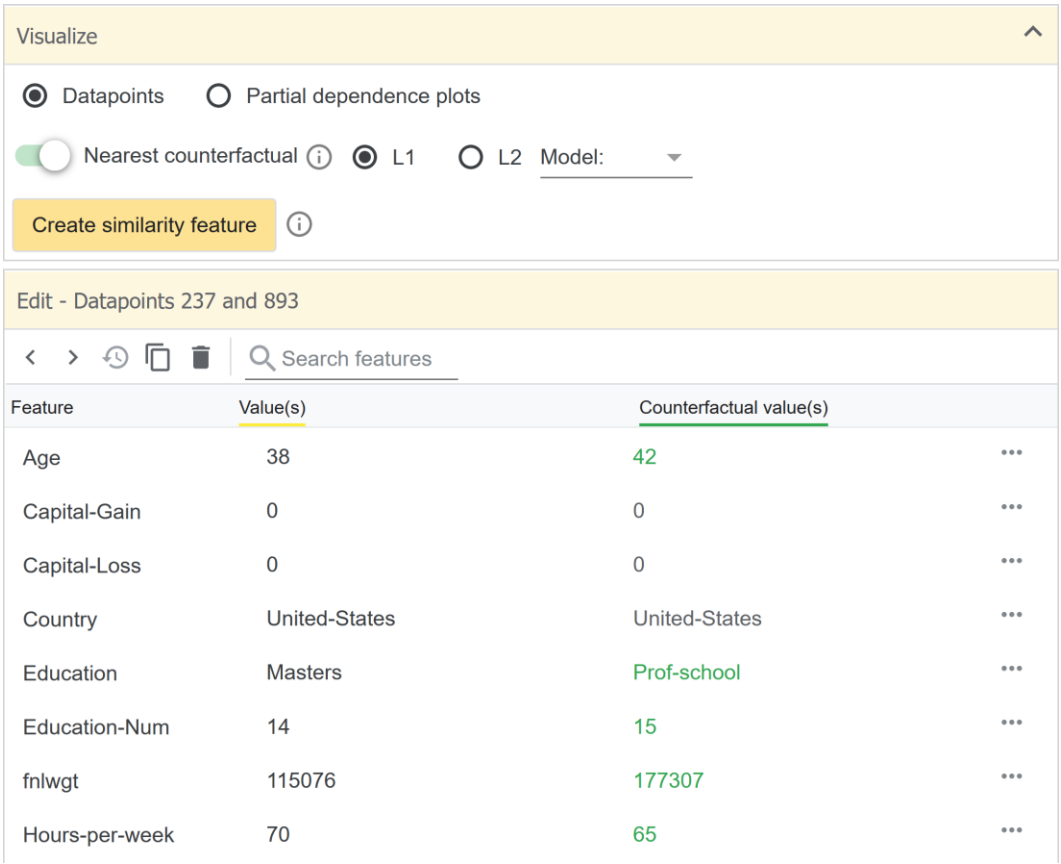
	Age	Purchasing	fnlwgt	Education	Education-Num	media	Occupation
0	39	shops	77516	Bachelors	13	live	Adm-clerical
1	50	markets	83311	Bachelors	13	DVD	Exec-managerial
2	38	malls	215646	HS-grad	9	social_networks	Handlers-cleaners
3	53	malls	234721	11th	7	DVD	Handlers-cleaners
4	28	malls	338409	Bachelors	13	DVD	Prof-specialty
...	...	...	...	...	...	...	...
10353	33	malls	126414	Bachelors	13	DVD	Other-service
10354	27	malls	43652	Bachelors	13	live	Adm-clerical
10355	47	supermarkets	227244	Bachelors	13	DVD	Protective-serv
10356	29	malls	160731	HS-grad	9	DVD	Craft-repair
10357	33	malls	287878	Some-college	10	DVD	Other-service




Visualize
Datapoints
Partial dependence plots
Nearest counterfactual
L1
L2
Model:
Create similarity feature
Edit - Datapoint 237
Search features
Feature
Value(s)
Age
38
Capital-Gain
0
Capital-Loss
0
Country
United-States
Education
Masters
Education-Num
14
Purchasing
ecommerce
reading
book
transport
car
traveling
action

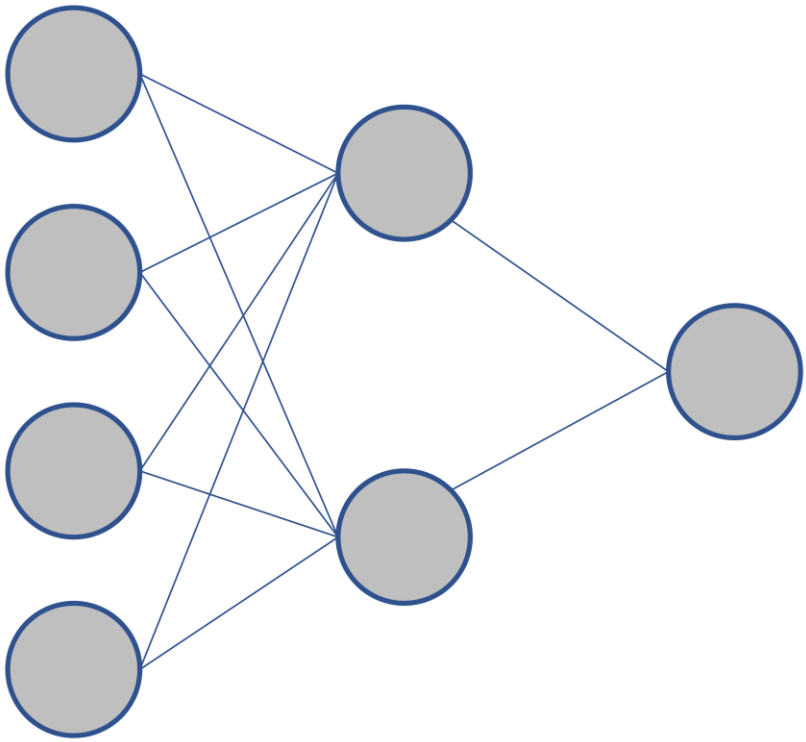

Visualize
Datapoints
Partial dependence plots
Nearest counterfactual
L1
L2
Model:
Create similarity feature


1
0.000078
0.00000893
Legend
Colors
by Inference label 1
Over 50K
Under 50K

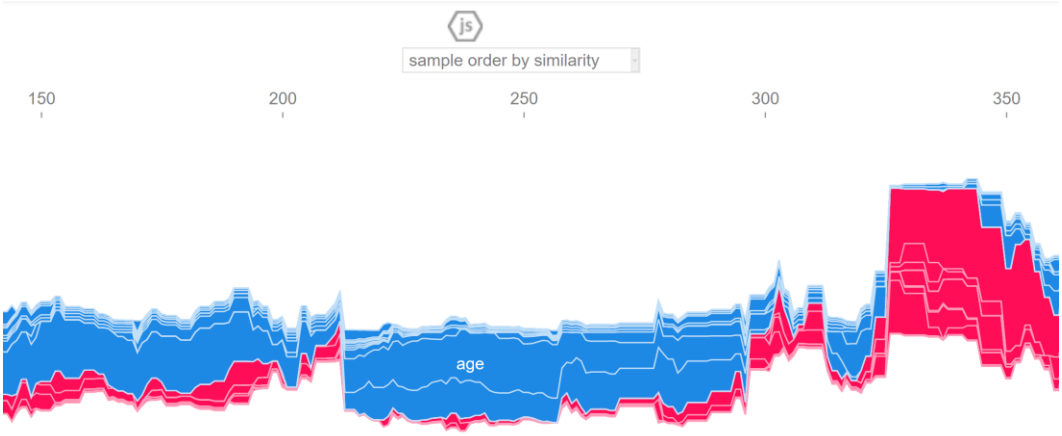

Visualize
Datapoints
Partial dependence plots
Nearest counterfactual
L1
L2
Model:
Create similarity feature
Edit - Datapoints 237 and 893
Search features
Feature
Value(s)
Counterfactual value(s)
Age 38 42
Capital-Gain 0 0
Capital-Loss 0 0
Country United-States United-States
Education Masters Prof-school
Education-Num 14 15
fnlwgt 115076 177307
Hours-per-week 70 65

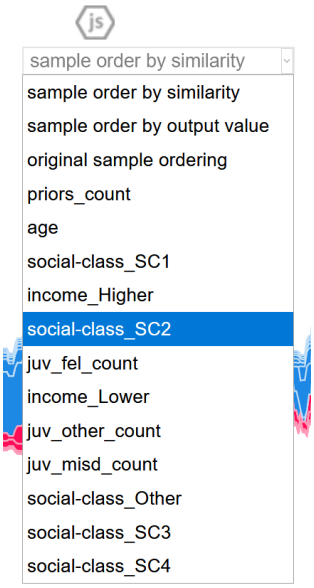
Chapter 06: AI Fairness with Google's What-If Tool (WIT)

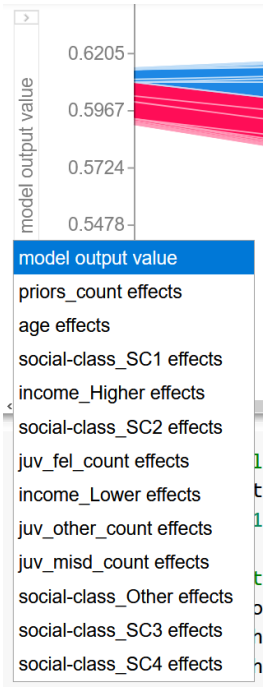
	age	juv_fel_count	juv_misd_count	juv_other_count	priors_count
0	69	0	0	0	0
1	69	0	0	0	0
3	34	0	0	0	0
4	24	0	0	1	4
5	24	0	0	1	4

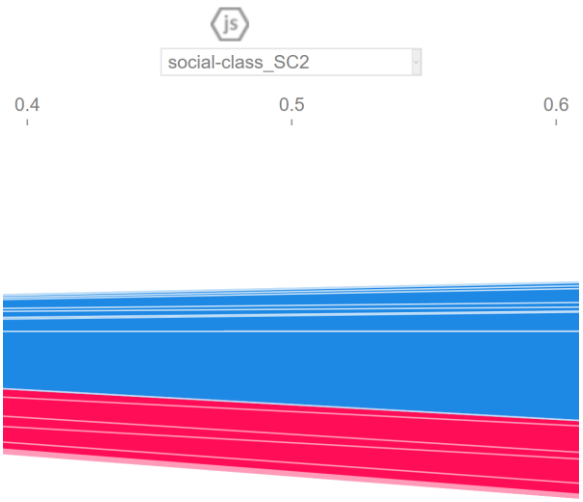


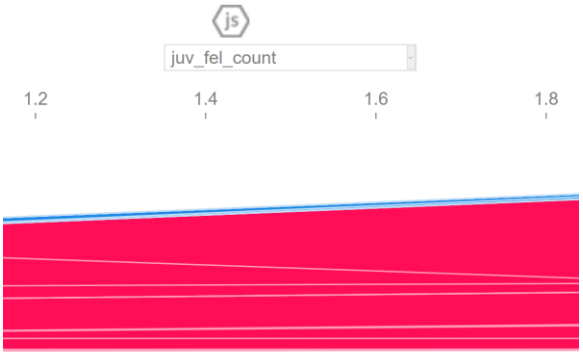
```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 200)               2600
_________________________________________________________________
dense_1 (Dense)              (None, 50)                10050
_________________________________________________________________
dense_2 (Dense)              (None, 25)                1275
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 26
=================================================================
Total params: 13,951
Trainable params: 13,951
Non-trainable params: 0
_________________________________________________________________
```



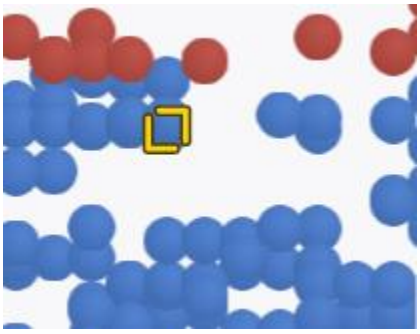

sample order by similarity
sample order by similarity
sample order by output value
original sample ordering
priors_count
age
social-class_SC1
income_Higher
social-class_SC2
juv_fel_count
income_Lower
juv_other_count
juv_misd_count
social-class_Other
social-class_SC3
social-class_SC4


model output value
0.6205
0.5967
0.5724
0.5478
model output value
priors_count effects
age effects
social-class_SC1 effects
income_Higher effects
social-class_SC2 effects
juv_fel_count effects
income_Lower effects
juv_other_count effects
juv_misd_count effects
social-class_Other effects
social-class_SC3 effects
social-class_SC4 effects


social-class_SC2
0.4
0.5
0.6

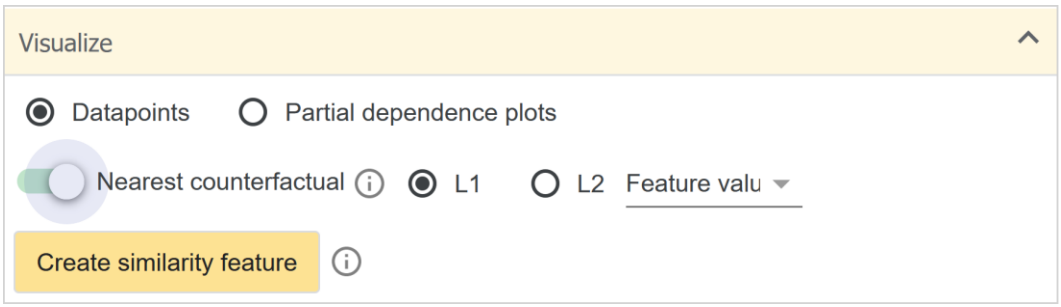

juv_fel_count
1.2
1.4
1.6
1.8

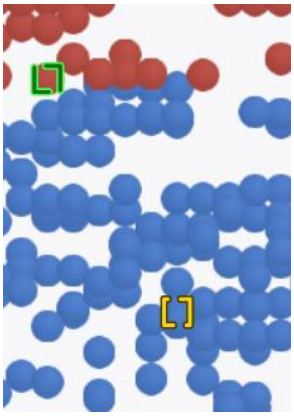
Binning | X-Axis
(none)
Binning | Y-Axis
(none)
Color By
Inference value
Label By
(default)
Scatter | X-Axis
(default)
Scatter | Y-Axis
Inference score
0.996
0.00162
Legend
Colors
by Inference value
0
1



Feature	Value(s)	Attribution value(s)
priors_count	7	0.2726
income_Higher	1	0.0955
social-class_SC2	1	0.0347
social-class_Other	0	0.0074
social-class_SC4	0	0
social-class_SC3	0	-0.0004
juv_misd_count	0	-0.0066
juv_other_count	0	-0.008
juv_fel_count	0	-0.0115

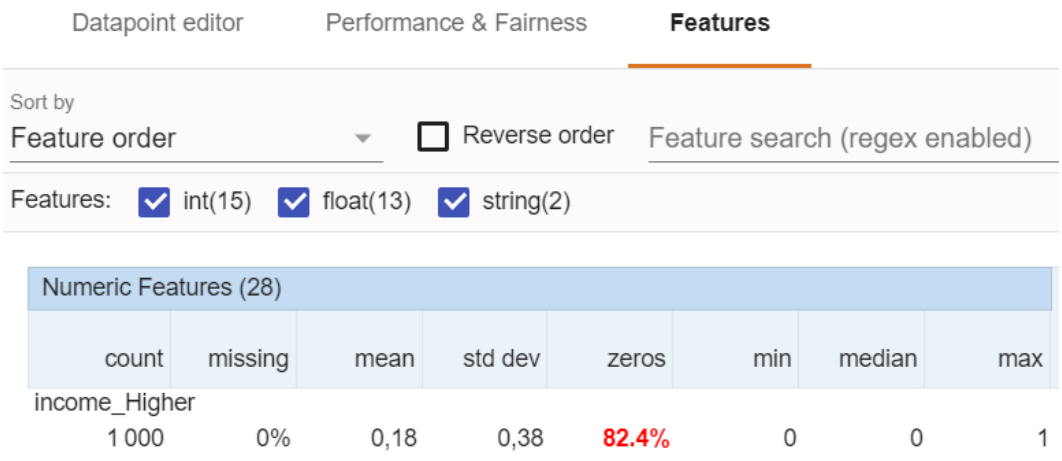
COMPASS_determination	0
recidivism_within_2_years	0

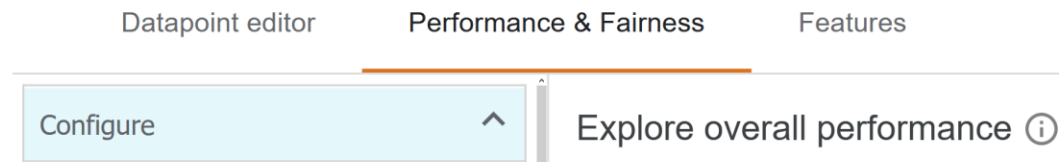


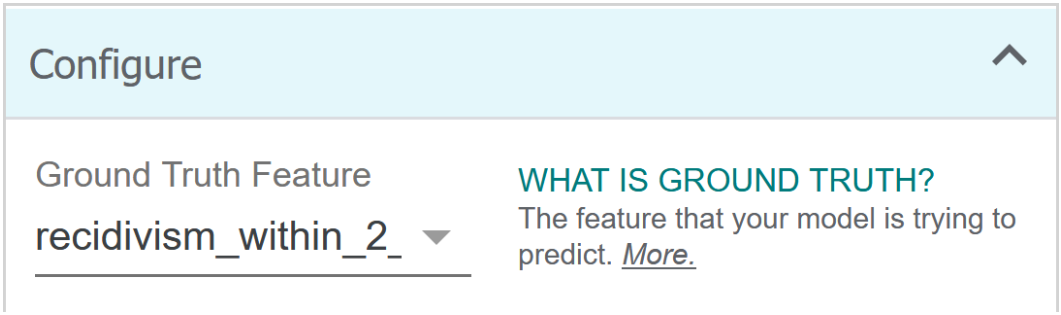


age	22	0.2266
social-class_SC1	1	0.0669
income_Lower	1	0.0174
social-class_Other	0	0.0077
social-class_SC4	0	0

income_Lower	0


Datapoint editor
Performance & Fairness
Features
Sort by
Feature order
Reverse order
Feature search (regex enabled)
Features:
int(15)
float(13)
string(2)
Numeric Features (28)
count
missing
mean
std dev
zeros
min
median
max
income_Higher
1 000
0%
0,18
0,38
82.4%
0
0
1


Datapoint editor
Performance & Fairness
Features
Configure
Explore overall performance


Configure
Ground Truth Feature
recidivism_within_2_
WHAT IS GROUND TRUTH?
The feature that your model is trying to predict. More.

Cost Ratio (FP/FN)

1

WHAT IS COST RATIO?

The cost of false positives relative to false negatives. Required for optimization. *More.*

Slice by

income_Higher

Buckets

2

WHAT DOES SLICING DO?

Shows the model's performance on datapoints grouped by each value of the selected feature.

Slice by (secondary)

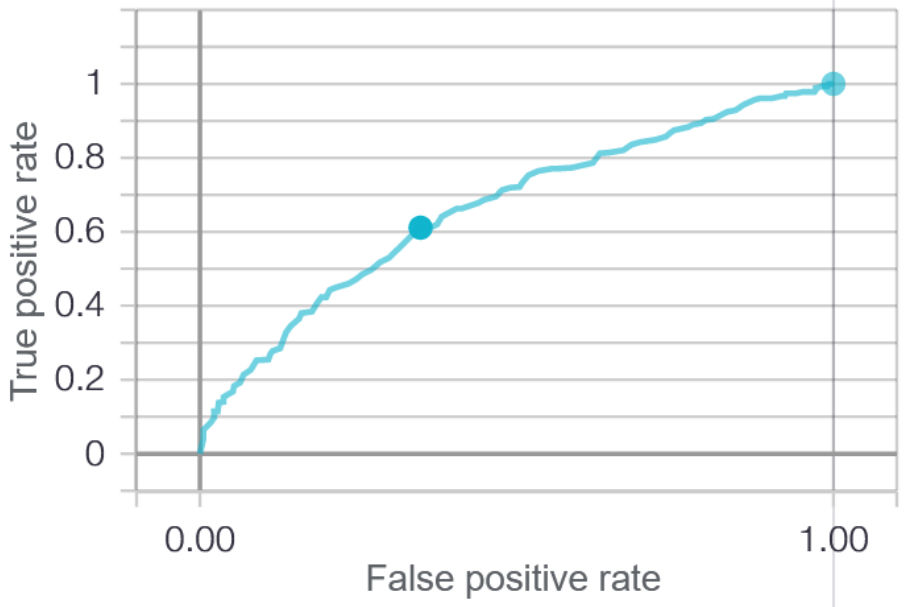
<none>

- (•) Custom thresholds
- ( ) Single threshold
- ( ) Demographic parity
- ( ) Equal opportunity
- ( ) Equal accuracy
- ( ) Group thresholds

Slice by

<none>

ROC curve (AUC: 0.66)

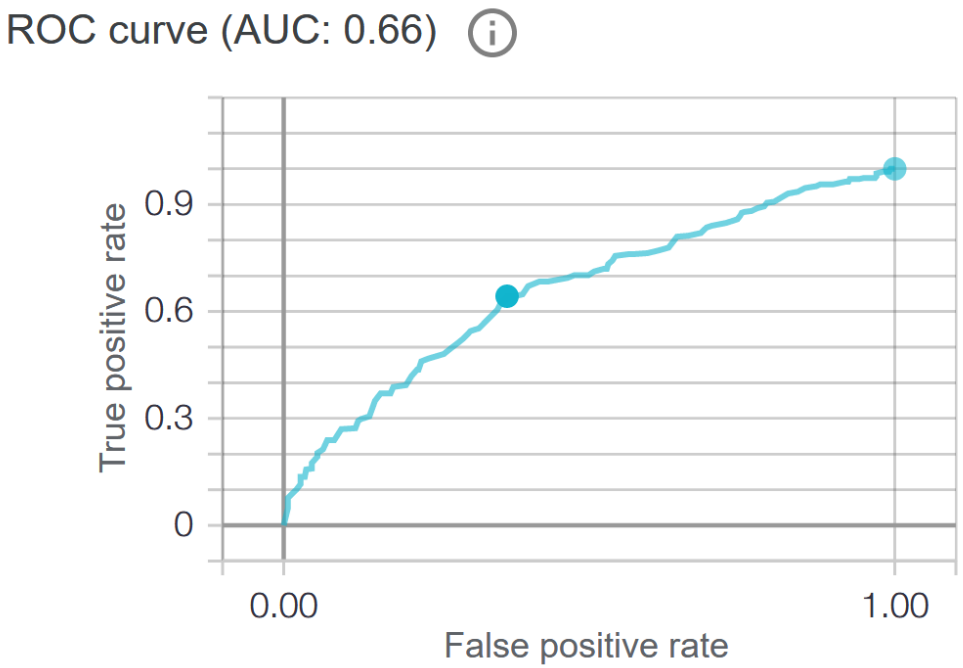


	Threshold	TPR	FPR
	0.50	0.7	0.5
	0.50	0.7	0.5

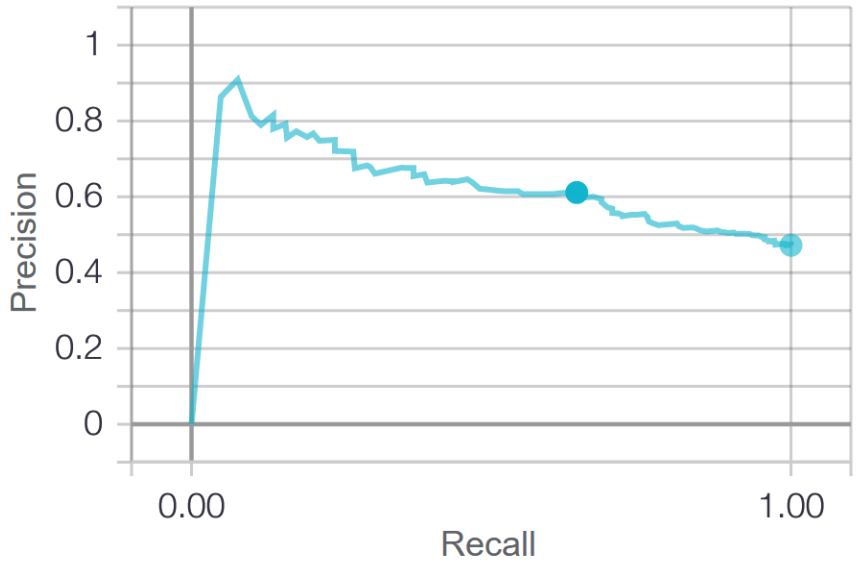
Custom thresholds for 2 values of income_Higher

Sort by: Count

Feature Value	Count	Threshold	False Positives (%)	False Negatives (%)	Accuracy (%)	F1
0	824	0.5	27.9	13.2	58.9	0.62
1	176	0.5	23.3	11.4	65.3	0.64



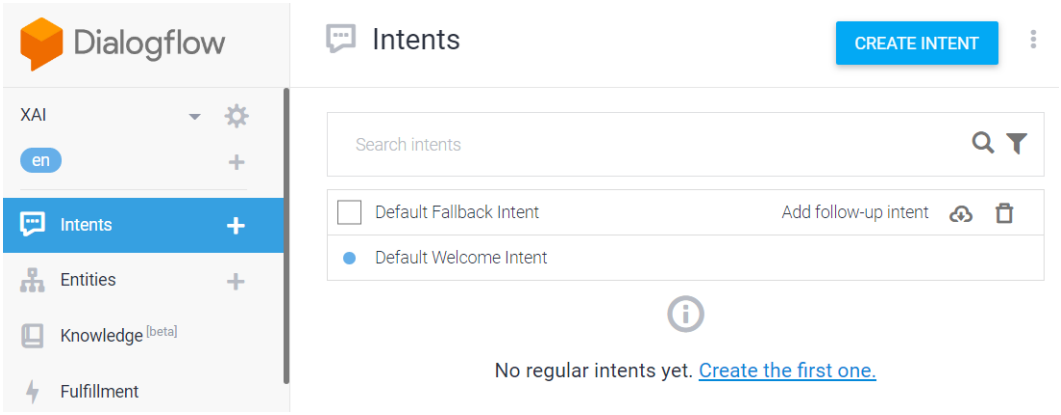
PR curve (AUC: 0.62)

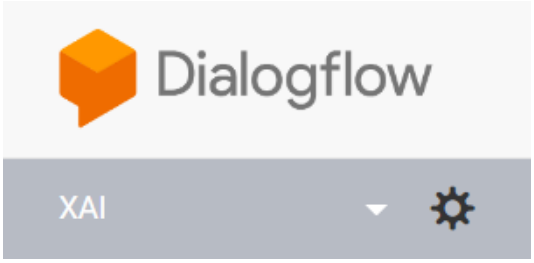


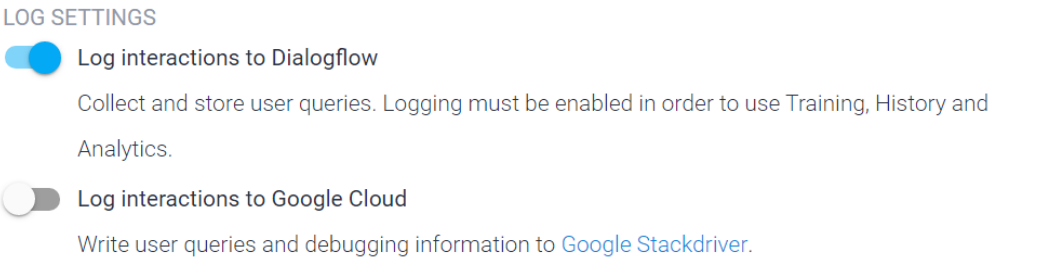
Confusion Matrix

	Predicted Yes		Predicted No		Total	
Actual Yes	30.3%	(250)	16.9%	(139)	47.2%	(389)
Actual No	19.3%	(159)	33.5%	(276)	52.8%	(435)
Total	49.6%	(409)	50.4%	(415)		

Chapter 07: A Python Client for Explainable AI Chatbots







General Languages

English – en

Try it now

Please use test console above to try a sentence.

See how it works in Google Assistant.

Agent

USER SAYS COPY CURL

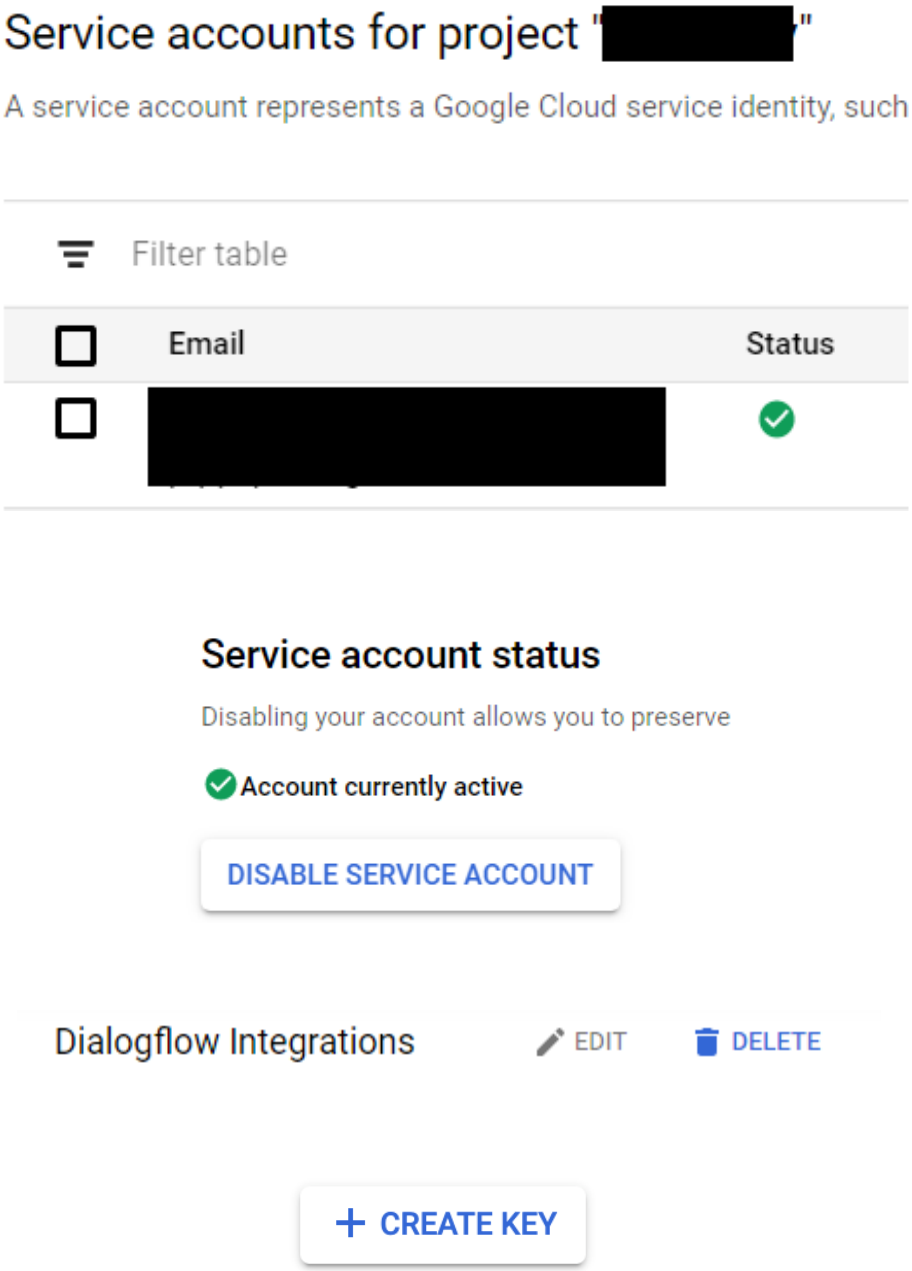
hi

DEFAULT RESPONSE

Hello! How can I help you?

GOOGLE PROJECT

Project ID	[redacted]
Service Account	[redacted]


Service accounts for project "[redacted]"
A service account represents a Google Cloud service identity, such
Filter table
Email
Status
[redacted]
Service account status
Disabling your account allows you to preserve
Account currently active
DISABLE SERVICE ACCOUNT
Dialogflow Integrations
EDIT
DELETE
+ CREATE KEY

Create private key for "Dialogflow Integrations"

Downloads a file that contains the private key. Store the file securely because this key can't be recovered if lost.

Key type

- (•) JSON
  Recommended
- ( ) P12
  For backward compatibility with code using the P12 format

CANCEL CREATE

Private key saved to your computer

⚠ [redacted] Learn more

Responses

DEFAULT +

Text Response	
1	Hi! How are you doing?
2	Hello! How can I help you?
3	Good day! What can I do for you today?
4	Greetings! How can I assist?

Training phrases

Add user expression

- just going to say hi
- heya
- hello hi
- howdy
- hey there
- hi there
- greetings
- hey

Dialogflow
XAI
en
Intents

CREATE INTENT

Intent name
SAVE

Training phrases

Train the intent with what your users will say

Provide examples of how users will express their intent in natural language. Adding numerous phrases with different variations and parameters will improve the accuracy of intent matching.
Learn more

ADD TRAINING PHRASES

Training phrases

Add user expression

I don't understand the Markov Decision Process

I don't understand the MDP

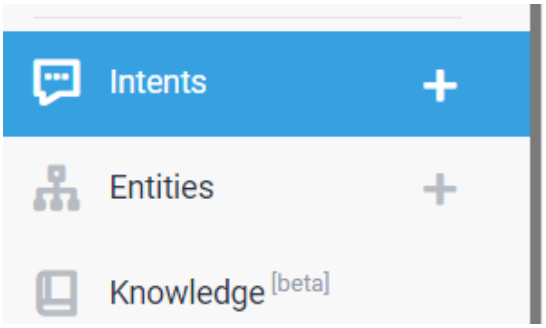
Bellman equation

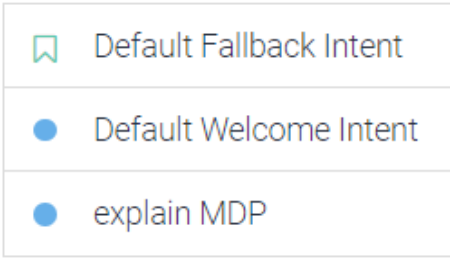
MDP Algorithm

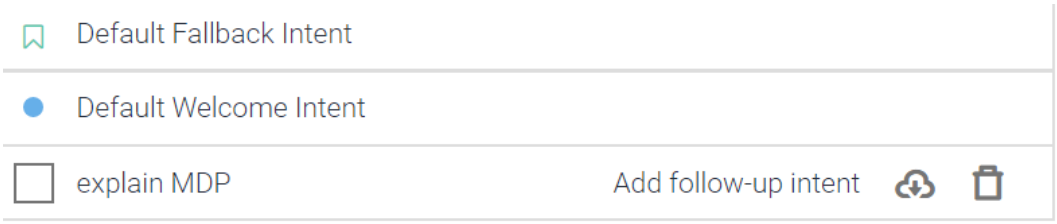
Explain this MDP algorithm

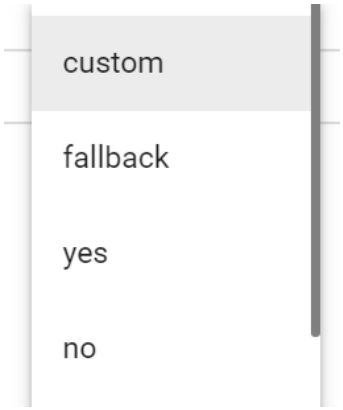
Text Response

1 The Markov Decision Process (MDP) uses information on the possible paths in your decision process that were provided by your team. It detects the best points you determined. They can be locations, intermediate decisions, sequences of words or anything you have decided. Would you like more information? Please answer yes to continue and no to end this intent dialog.



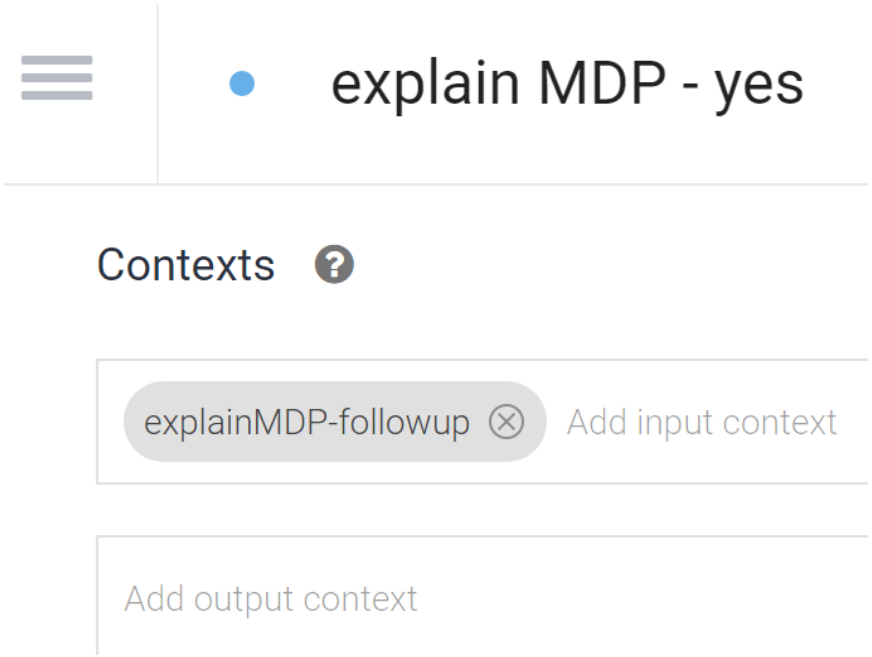






explain MDP

explain MDP - yes


explain MDP - yes
Contexts
explainMDP-followup
Add input context
Add output context

Training phrases

Add user expression

- yes
- okay I will
- why not
- yes that's alright
- yes I do

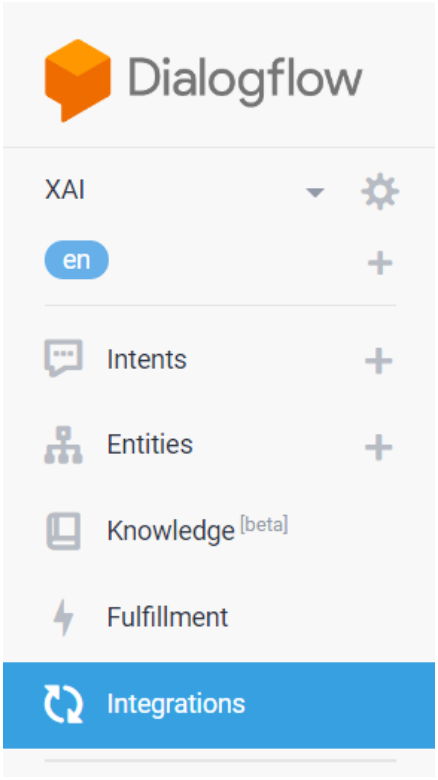
Text Response

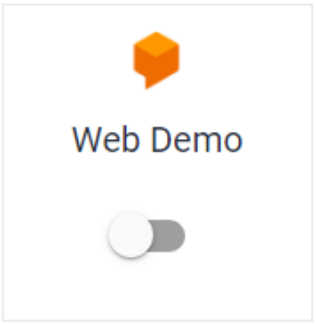
1 If this dialog was triggered, there must be external data linked to your MDP process. In this application, it is Work In Process (WIP). A alert has been triggered because a threshold has been exceeded. This threshold is not acceptable for your application. We recommend you interrupt the machine learning program and investigate the reasons of these high values.

ADD RESPONSES

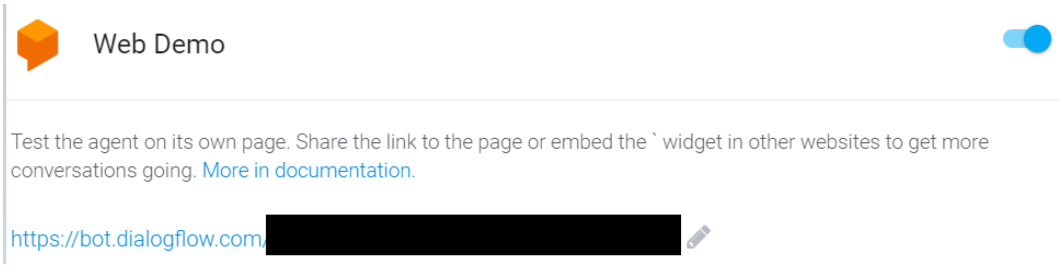
Set this intent as end of conversation

XAI Chatbot



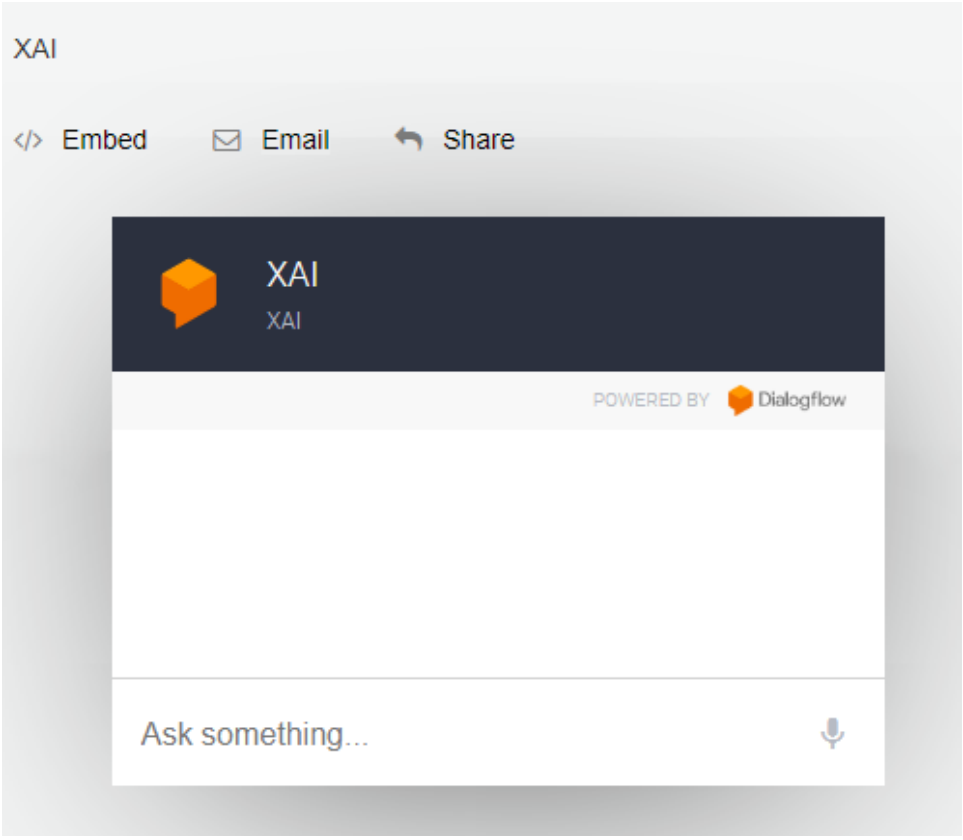

Web Demo

Web Demo


Web Demo
Test the agent on its own page. Share the link to the page or embed the ` widget in other websites to get more conversations going. More in documentation.
https://bot.dialogflow.com/[redacted]

Add this agent to your website by copying the code below:

```
<iframe
    allow="microphone;"
    width="350"
    height="430"
    src="https://console.dialogflow.com/api-client/demo/embedded/[redacted]
220">
</iframe>
```



Ask something...

hello

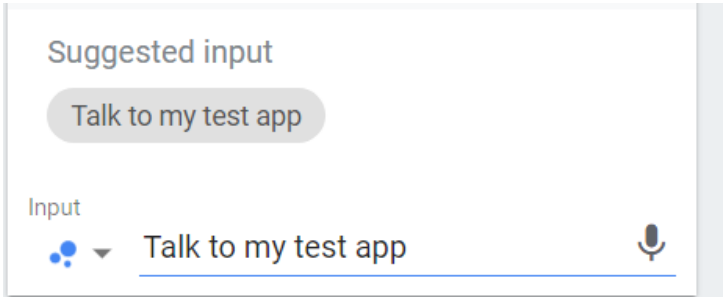
Hello! How can I help you?

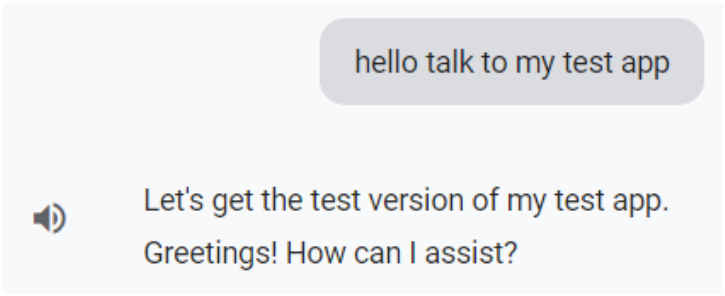
Try it now

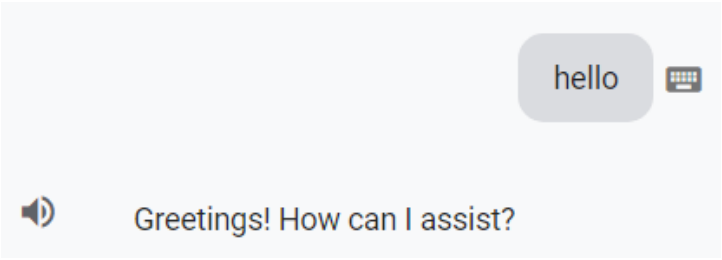
Please use test console above to try a sentence.

See how it works in Google Assistant.

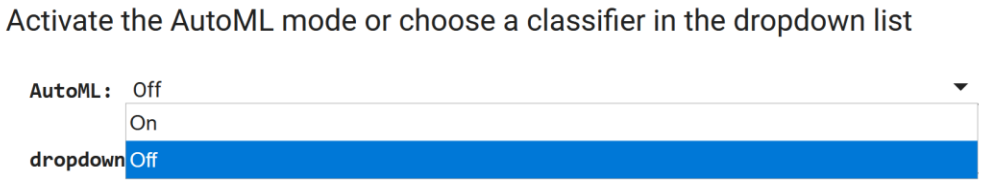
Updating Actions on Google...

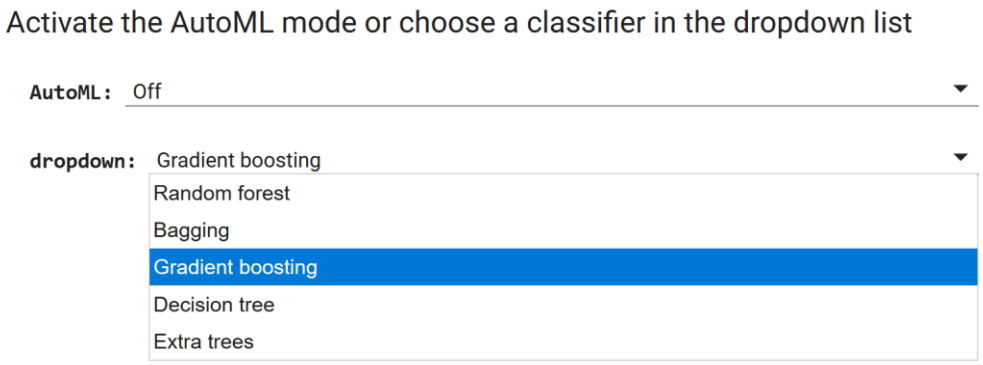


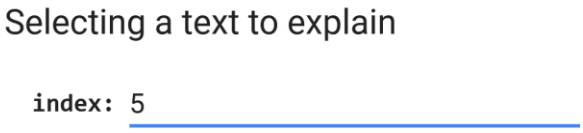


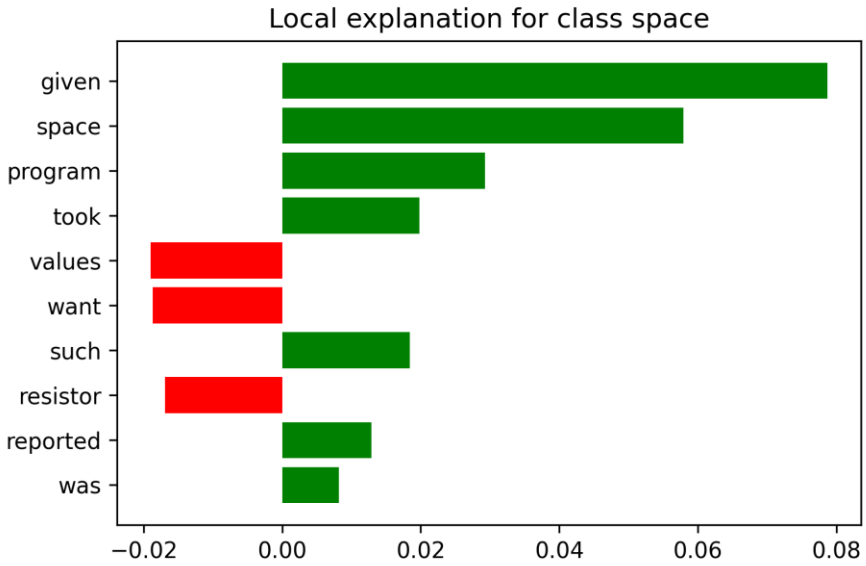

hello
Greetings! How can I assist?

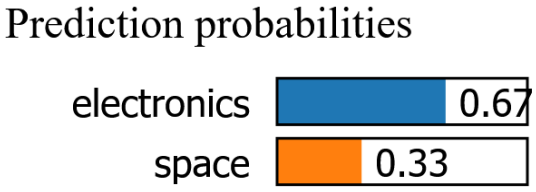
Chapter 08: Local Interpretable Model-Agnostic Explanations (LIME)

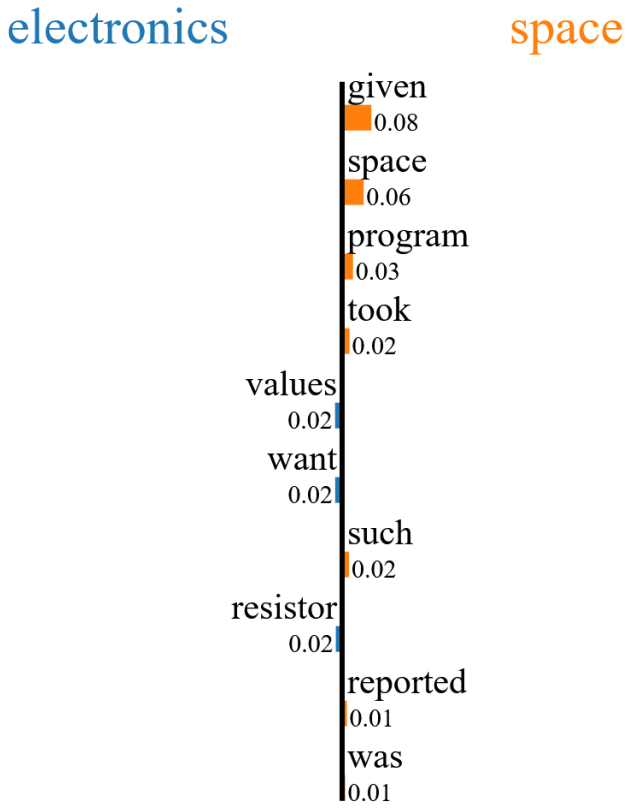







Local explanation for class space
given
space
program
took
values
want
such
resistor
reported
was
−0.02
0.00
0.02
0.04
0.06
0.08

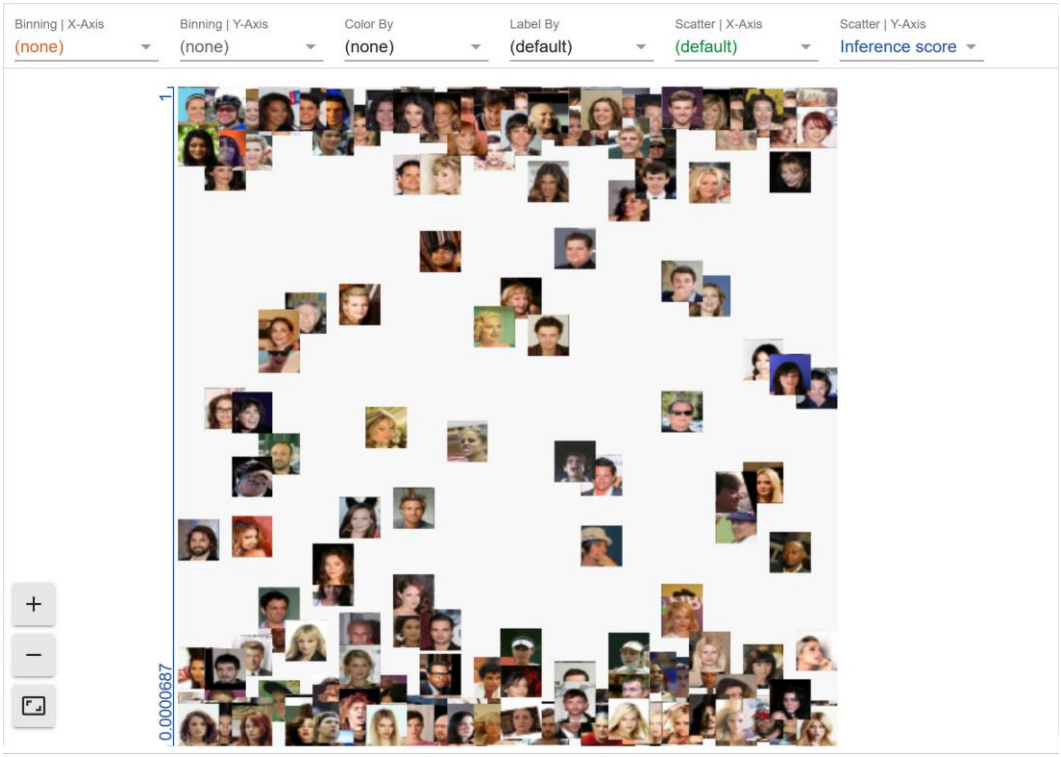

Prediction probabilities
electronics
0.67
space
0.33

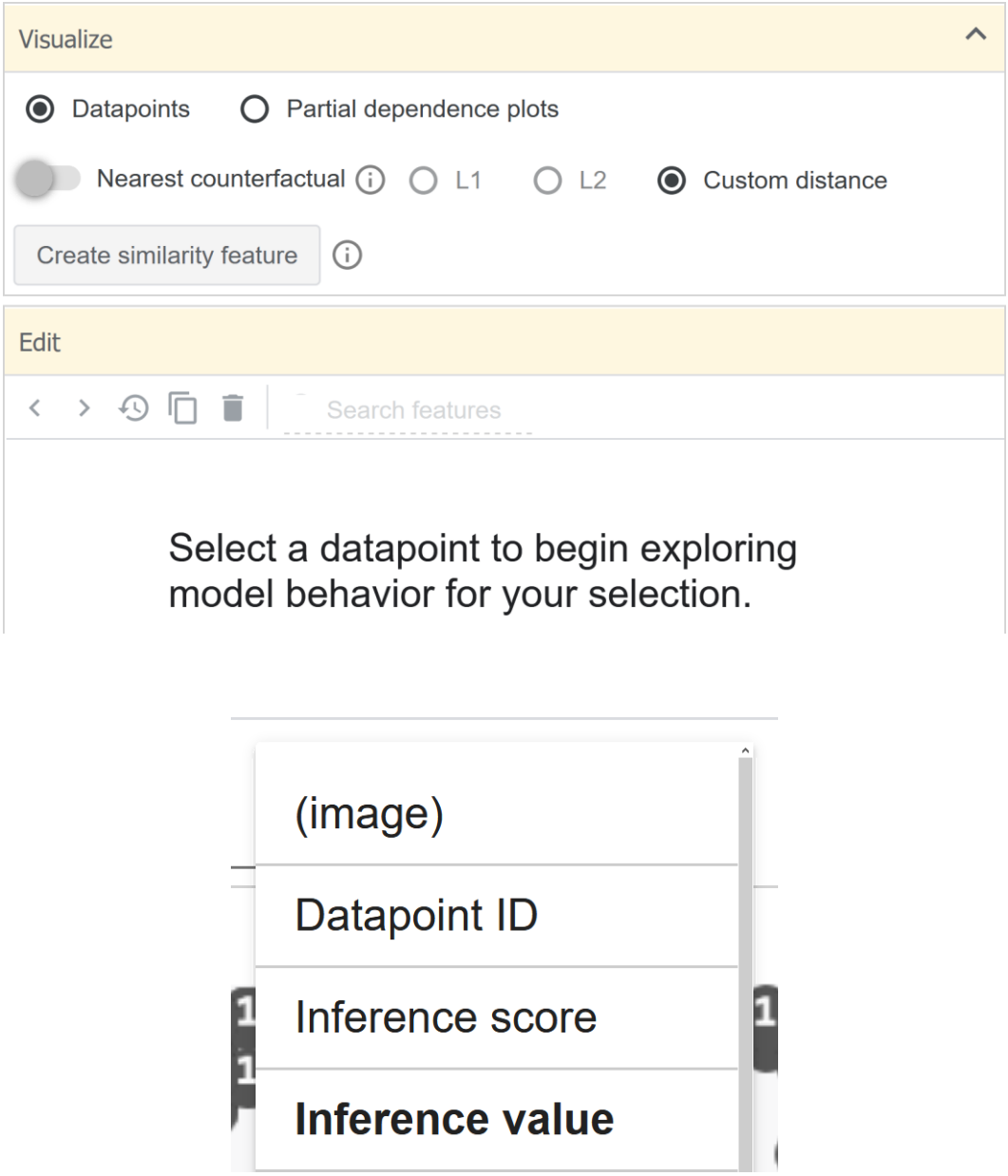


|Is there a PD program out there that will work out how best to make up such |a resistance, given fixed resistors of the standard 12 values per decade?.(1, |1.2,1.5,1.8,2.2,3.3 etc). It is a common enough problem, yet I cant |recall seing a program that tells that Rx+Ry//Rz gives Rq, starting with |q and finding prefered values x,y and z.

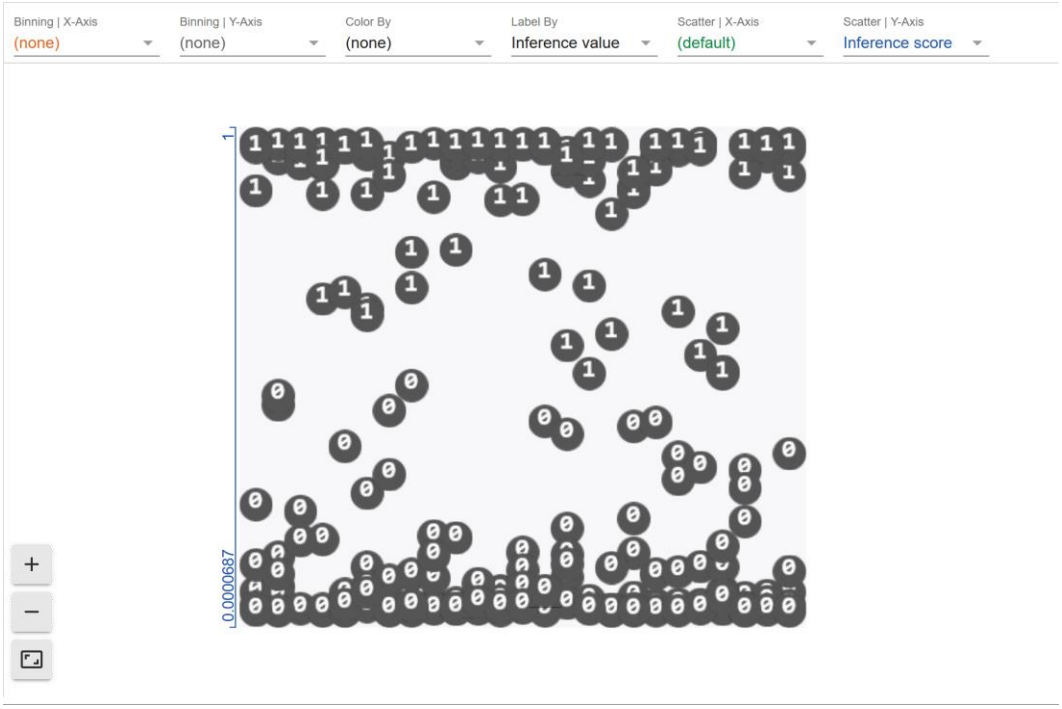
I once wrote such a program (in BBC basic...) It was very crude, and took around 5 seconds to do an exhaustive search (with a small amount of intelligence), and told you the best combination |Rq and the best below Rq.

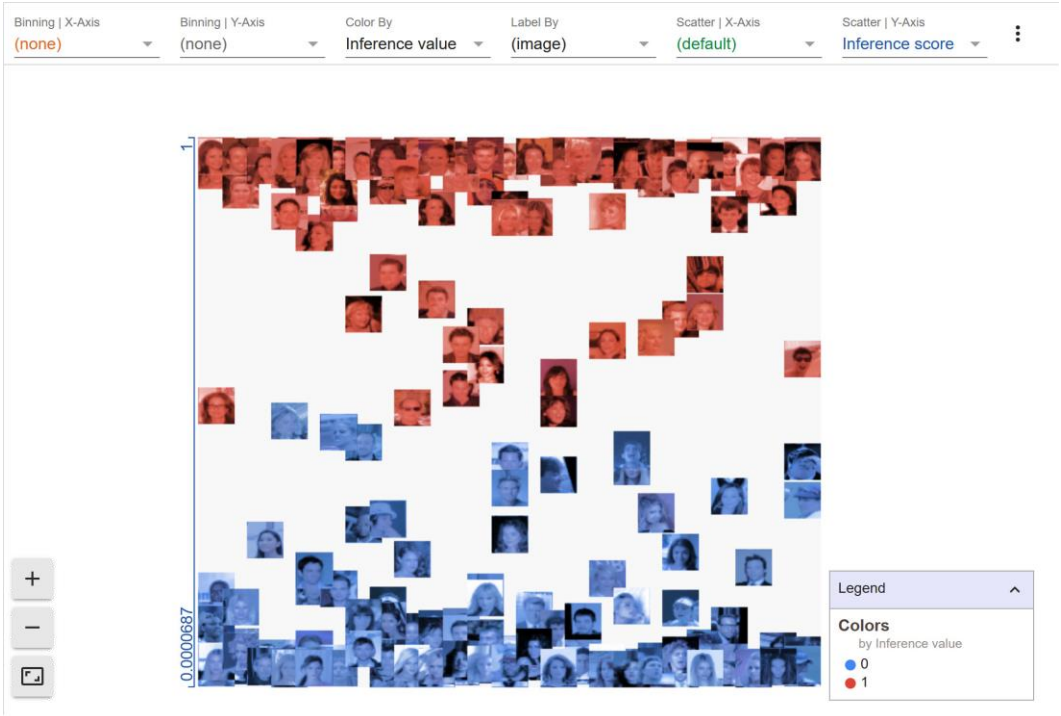
Chapter 09: The Counterfactual Explanations Method

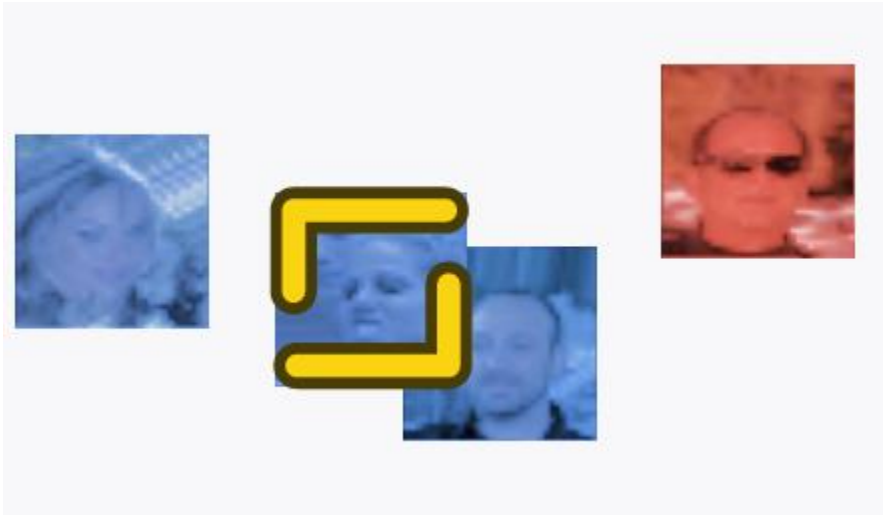



Visualize
Datapoints
Partial dependence plots
Nearest counterfactual
L1
L2
Custom distance
Create similarity feature
Edit
Search features
Select a datapoint to begin exploring model behavior for your selection.

(image)
Datapoint ID
Inference score
Inference value


Binning | X-Axis
(none)
Binning | Y-Axis
(none)
Color By
(none)
Label By
Inference value
Scatter | X-Axis
(default)
Scatter | Y-Axis
Inference score
1
0.0000687


Binning | X-Axis
(none)
Binning | Y-Axis
(none)
Color By
Inference value
Label By
(image)
Scatter | X-Axis
(default)
Scatter | Y-Axis
Inference score
1
0.0000687
Legend
Colors
by Inference value
0
1



Edit - Datapoint 226

Search features

Feature	Value(s)	
image/encoded		•••
5_o_Clock_Shadow	0	•••
Arched_Eyebrows	0	•••

Visualize

Datapoints Partial dependence plots

Nearest counterfactual L1 L2 Custom distance

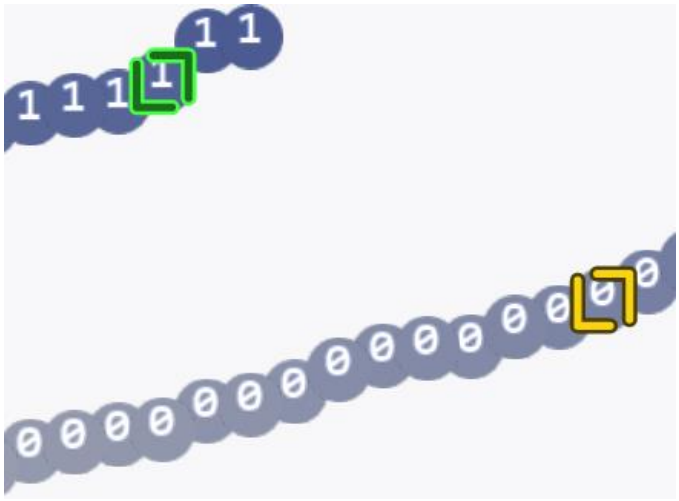
Create similarity feature

Edit - Datapoints 226 and 138

Search features

Feature	Value(s)	Counterfactual value(s)	
image/encoded			•••
5_o_Clock_Shadow	0	0	•••
Arched_Eyebrows	0	0	•••

Sideburns	0	0	•••
Smiling	0	0	•••
Straight_Hair	0	0	•••



Edit - Datapoints 11 and 57

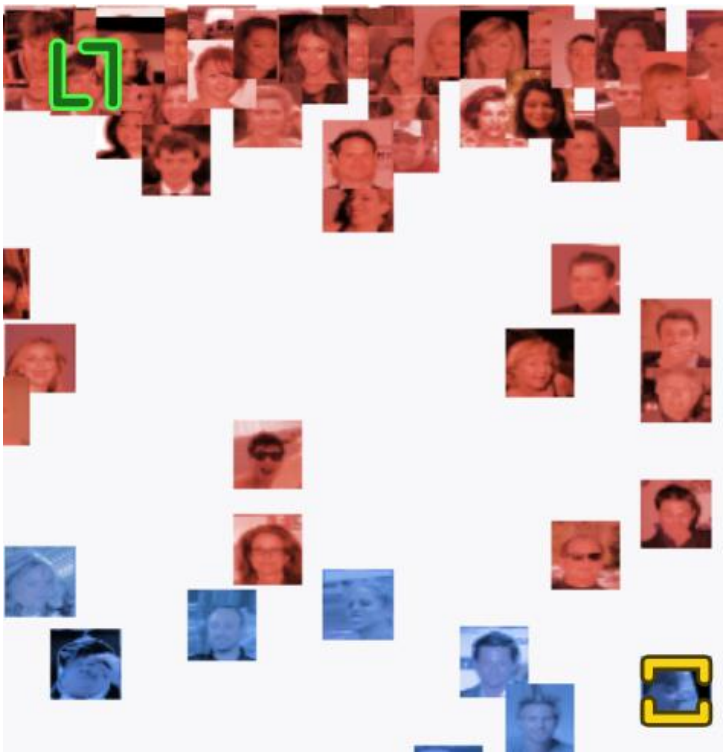
Search features

Feature	Value(s)	Counterfactual value(s)	
image/encoded			•••

Edit - Datapoints 164 and 1

Search features

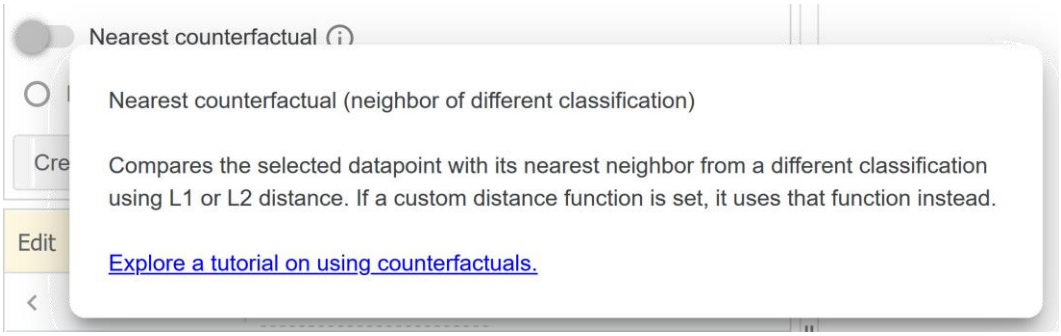
Feature	Value(s)	Counterfactual value(s)	
image/encoded			•••
5_o_Clock_Shadow	0	0	•••

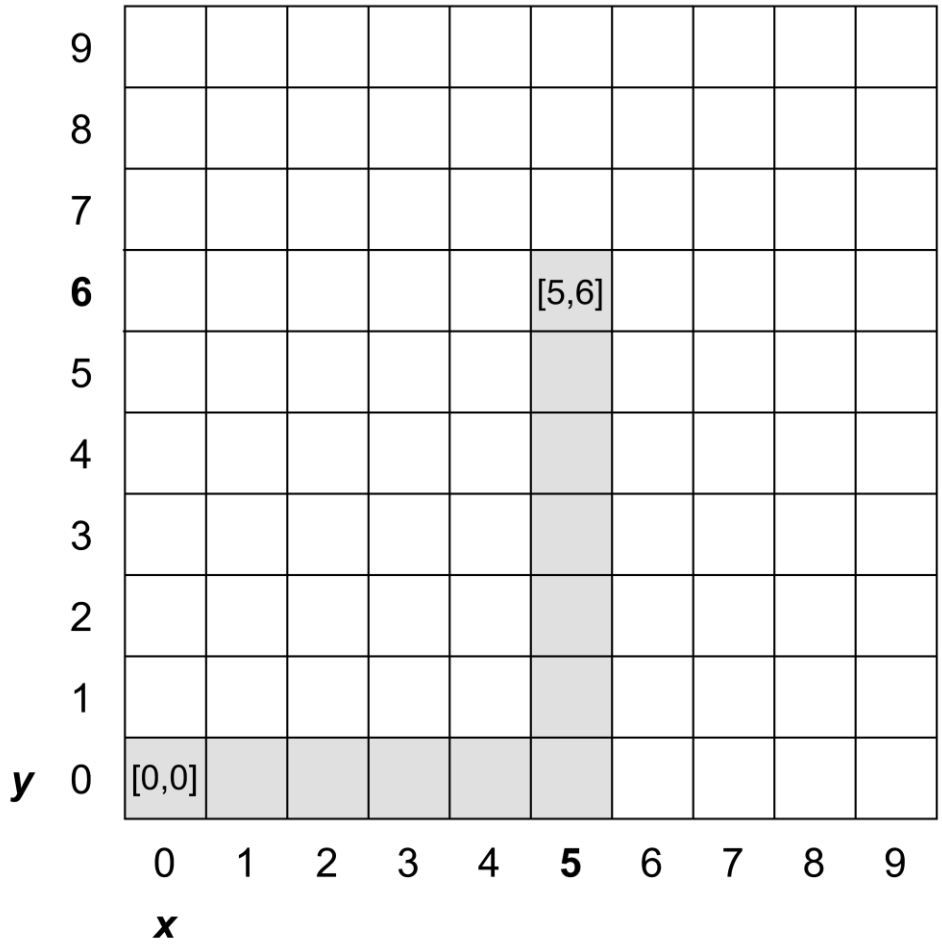


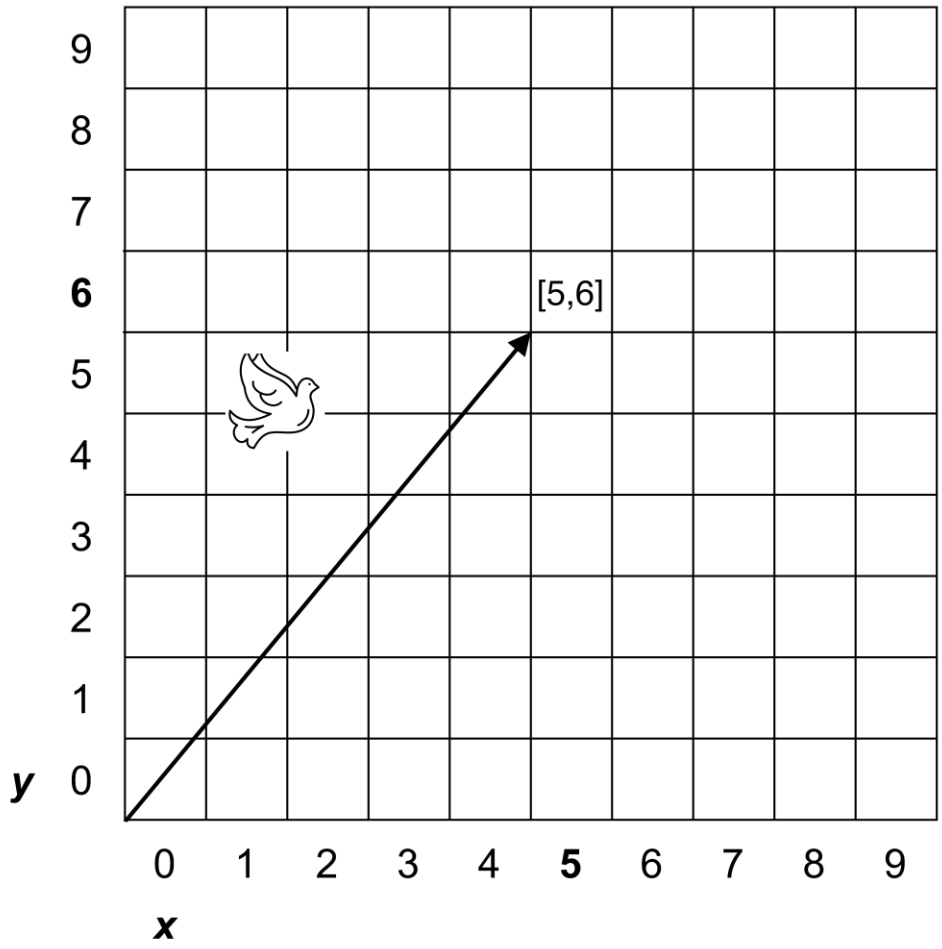
Edit - Datapoints 90 and 68

Search features

Feature	Value(s)	Counterfactual value(s)
image/encoded		
5_o_Clock_Shadow	0	0


Nearest counterfactual
Nearest counterfactual (neighbor of different classification)
Compares the selected datapoint with its nearest neighbor from a different classification using L1 or L2 distance. If a custom distance function is set, it uses that function instead.
Explore a tutorial on using counterfactuals.
Edit

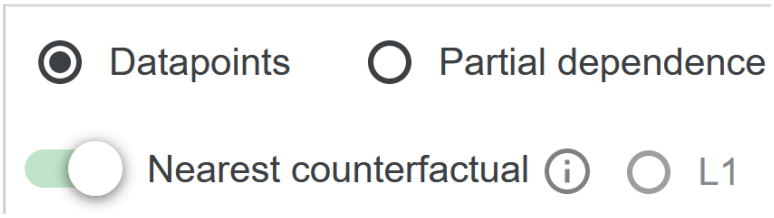

9
8
7
6
5
4
3
2
1
0
y
[5,6]
[0,0]
0 1 2 3 4 5 6 7 8 9
x



Invoke What-If Tool for the data and model

num_datapoints: 250

tool_height_in_px: 750



Edit - Datapoints 51 and 84

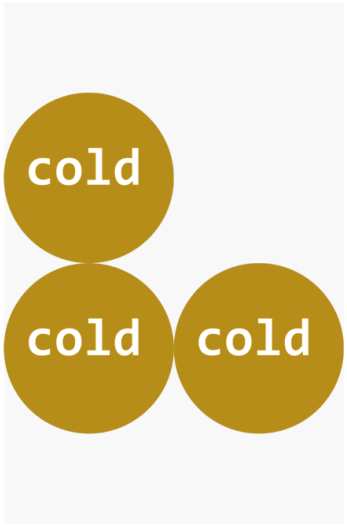
Search features

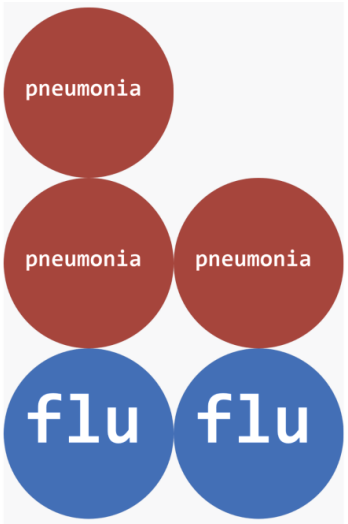
Feature	Value(s)	Counterfactual value(s)	
image/encoded			•••
5_o_Clock_Shadow	1	0	•••

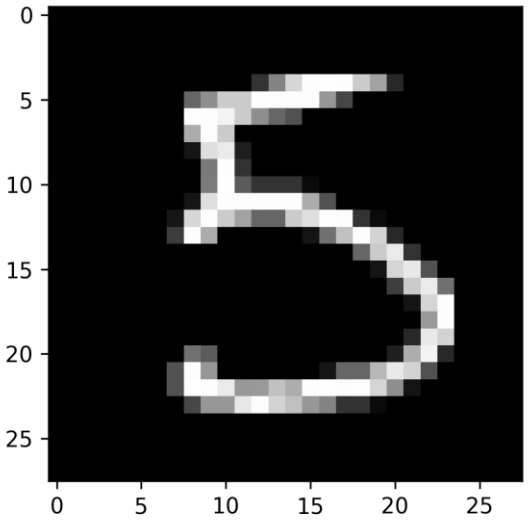
Infer - Datapoints 51 and 84

Predict

Chapter 10: Contrastive XAI

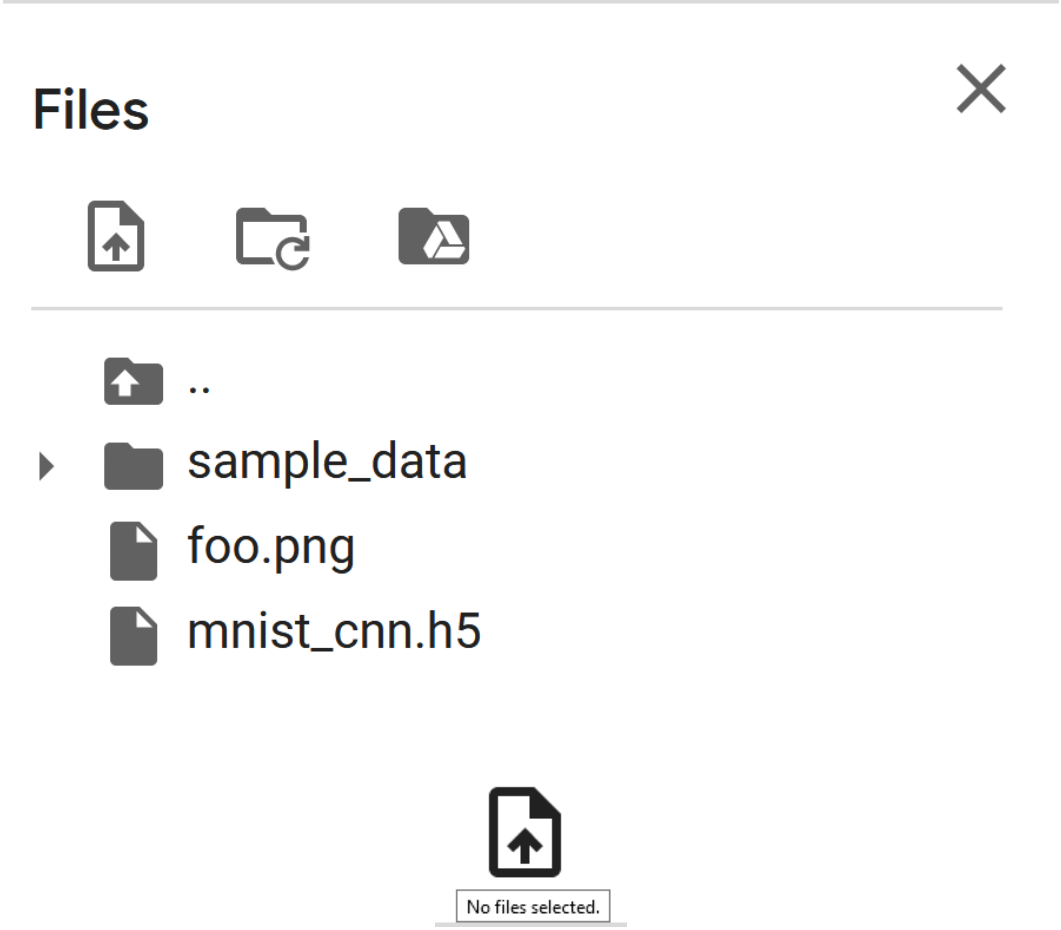






```
Model: "model"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 28, 28, 1)]       0
_________________________________________________________________
conv2d (Conv2D)              (None, 28, 28, 64)        320
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 14, 14, 64)        0
_________________________________________________________________
dropout (Dropout)            (None, 14, 14, 64)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 14, 14, 32)        8224
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 7, 7, 32)          0
_________________________________________________________________
dropout_1 (Dropout)          (None, 7, 7, 32)          0
_________________________________________________________________
flatten (Flatten)            (None, 1568)              0
_________________________________________________________________
dense (Dense)                (None, 256)               401664
_________________________________________________________________
dropout_2 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_1 (Dense)              (None, 10)                2570
=================================================================
```

train_cnn: yes

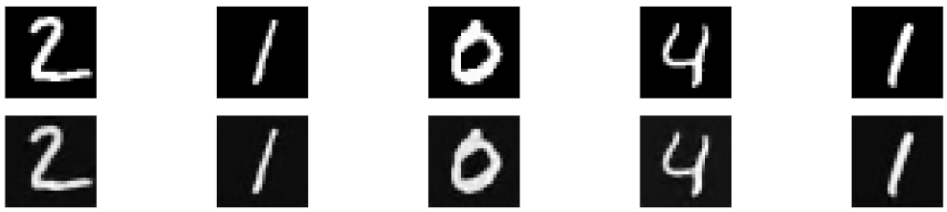

Files
..
sample_data
foo.png
mnist_cnn.h5
No files selected.

```
Model: "model"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 28, 28, 1)]       0
_________________________________________________________________
conv2d (Conv2D)              (None, 28, 28, 64)        320
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 14, 14, 64)        0
_________________________________________________________________
dropout (Dropout)            (None, 14, 14, 64)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 14, 14, 32)        8224
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 7, 7, 32)          0
_________________________________________________________________
dropout_1 (Dropout)          (None, 7, 7, 32)          0
_________________________________________________________________
flatten (Flatten)            (None, 1568)              0
_________________________________________________________________
dense (Dense)                (None, 256)               401664
_________________________________________________________________
dropout_2 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_1 (Dense)              (None, 10)                2570
=================================================================
Total params: 412,778
Trainable params: 412,778
Non-trainable params: 0
_________________________________________________________________
Test accuracy:  0.9867
```

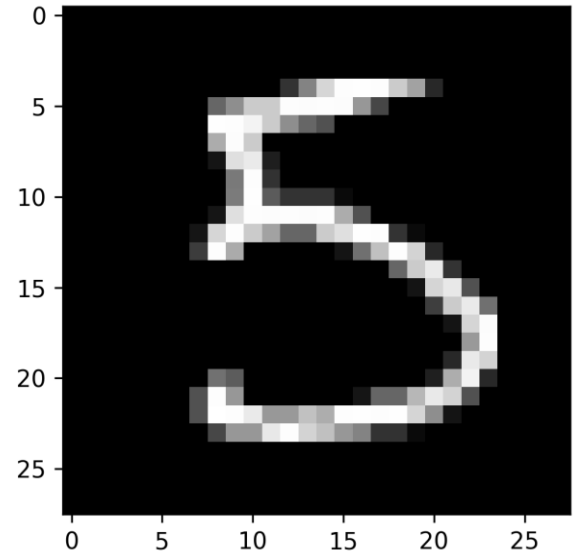
train_auto_encoder: yes

mnist_ae.h5

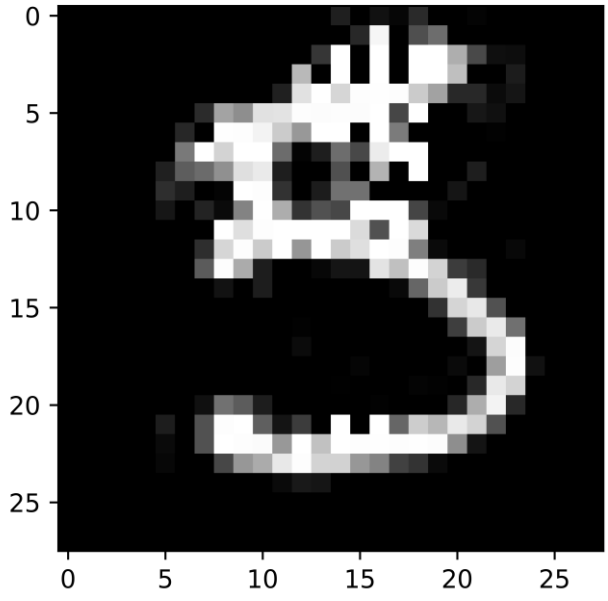
```
Model: "model_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         [(None, 28, 28, 1)]       0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 28, 28, 16)        160
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 28, 28, 16)        2320
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 14, 14, 16)        0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 14, 14, 1)         145
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 14, 14, 16)        160
_________________________________________________________________
up_sampling2d (UpSampling2D) (None, 28, 28, 16)        0
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 28, 28, 16)        2320
_________________________________________________________________
conv2d_7 (Conv2D)            (None, 28, 28, 1)         145
=================================================================
Total params: 5,250
Trainable params: 5,250
Non-trainable params: 0
_________________________________________________________________
```



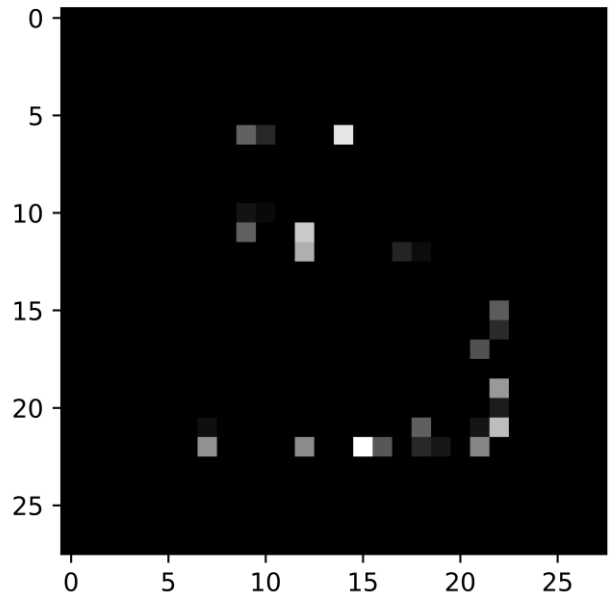
8 3



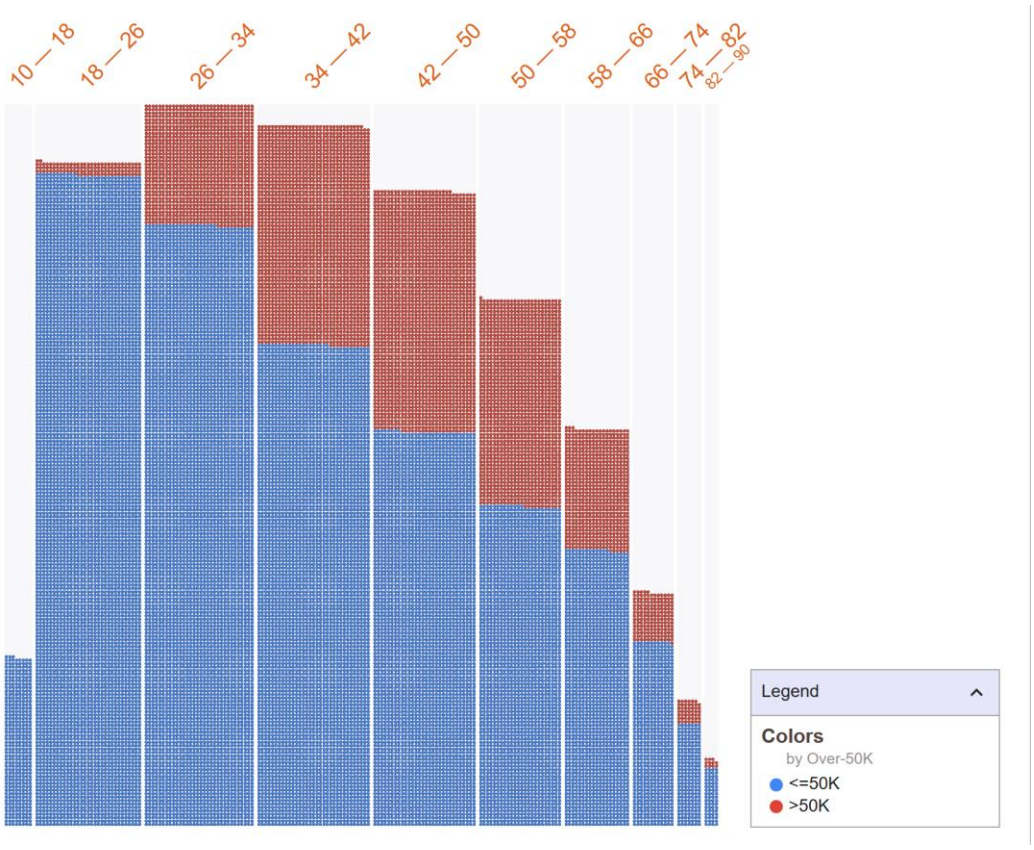

```
Pertinent negative prediction: 3
```

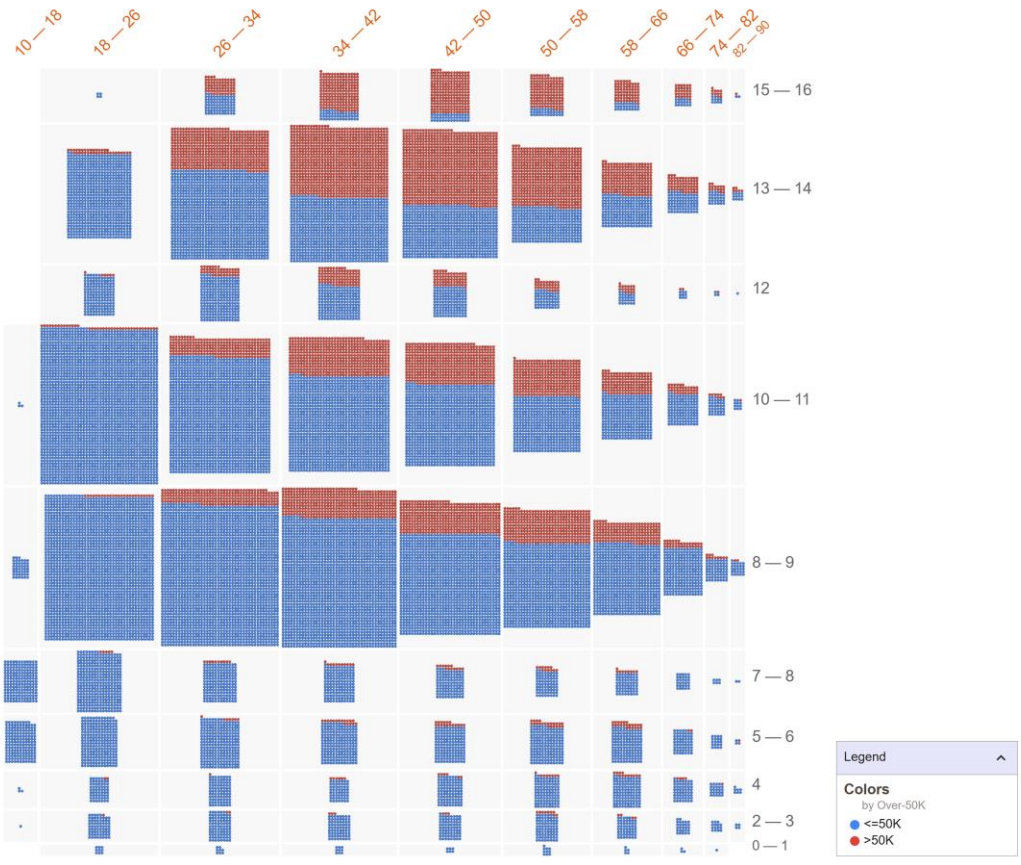


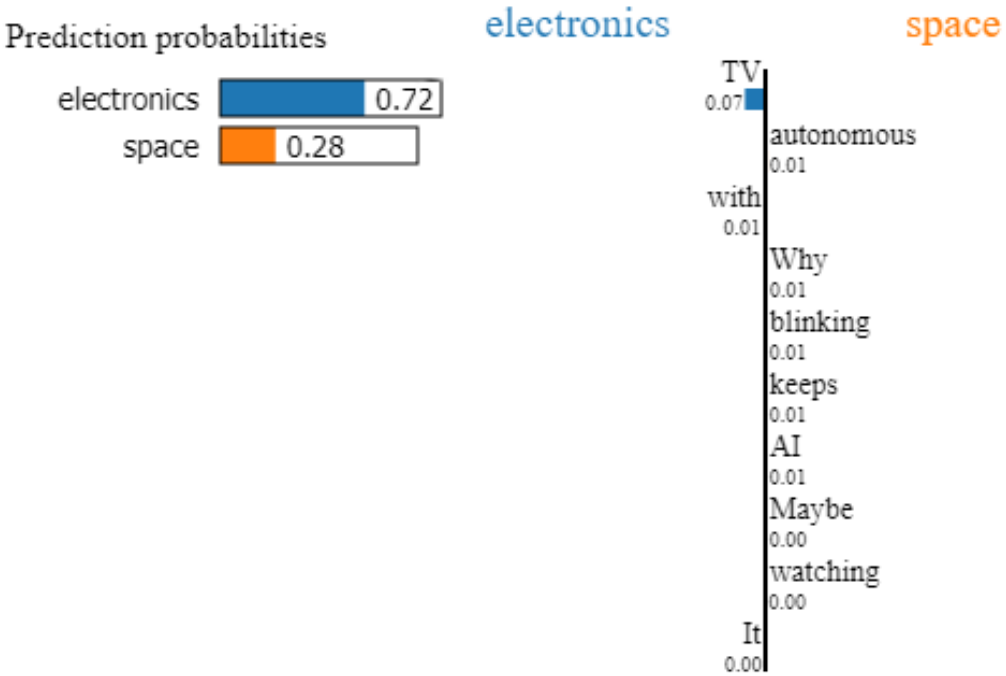
Pertinent positive prediction: 5



Chapter 11: Anchors XAI

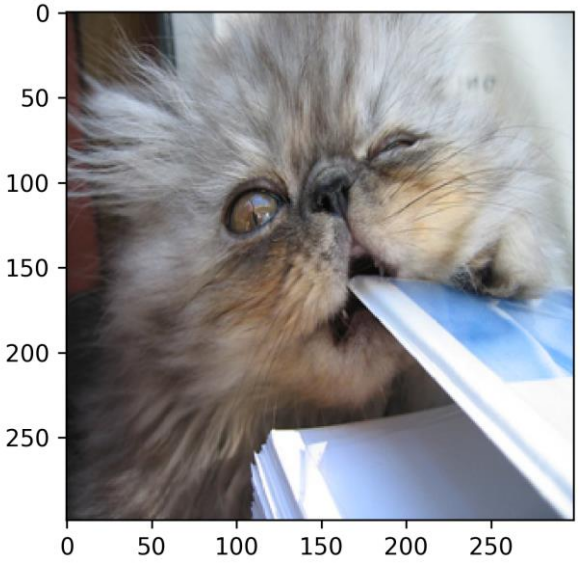


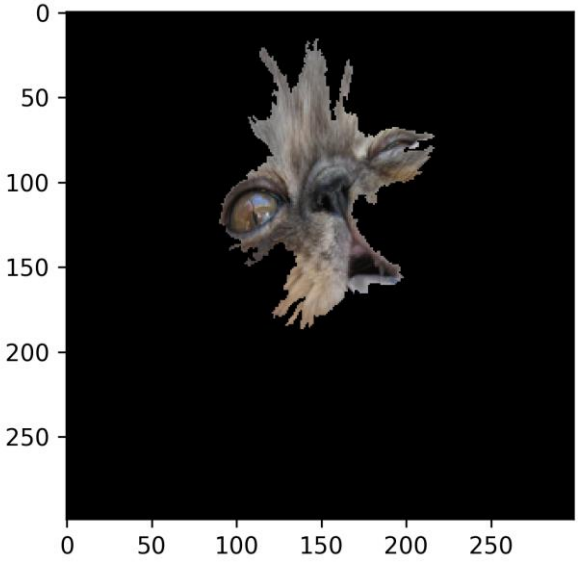

10 — 18
18 — 26
26 — 34
34 — 42
42 — 50
50 — 58
58 — 66
66 — 74
74 — 82
82 — 90
15 — 16
13 — 14
12
10 — 11
8 — 9
7 — 8
5 — 6
4
2 — 3
0 — 1
Legend
Colors
by Over-50K
<=50K
>50K

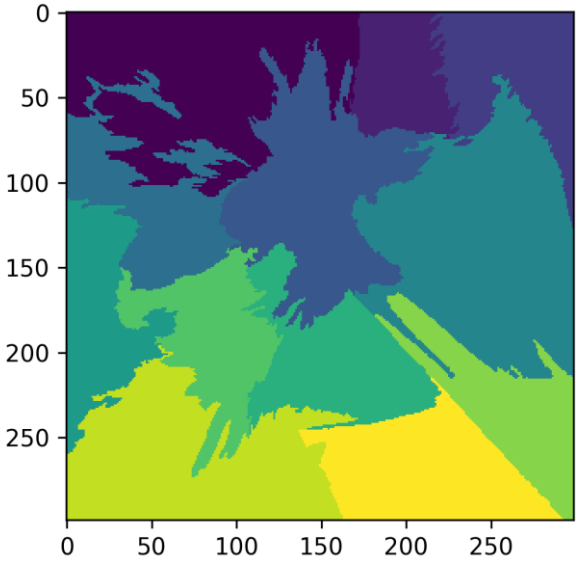


Download image from ImageNet

category: Persian cat

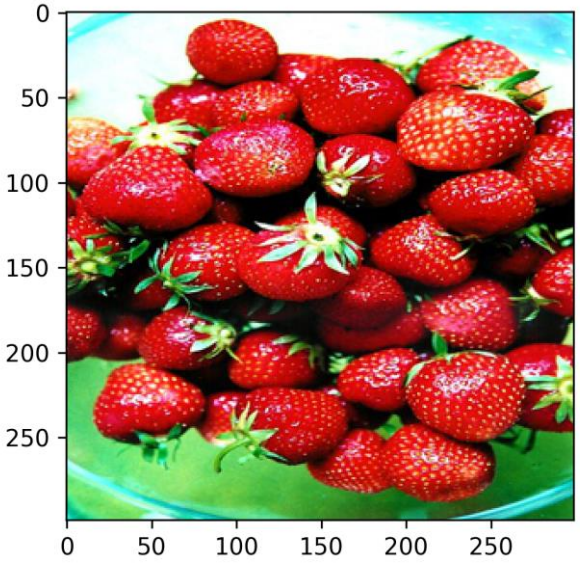

0
50
100
150
200
250
0
50
100
150
200
250

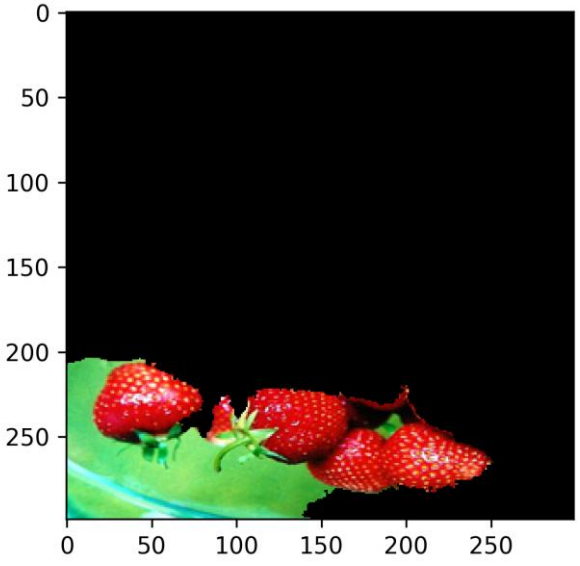

0
50
100
150
200
250
0
50
100
150
200
250



Download image from ImageNet

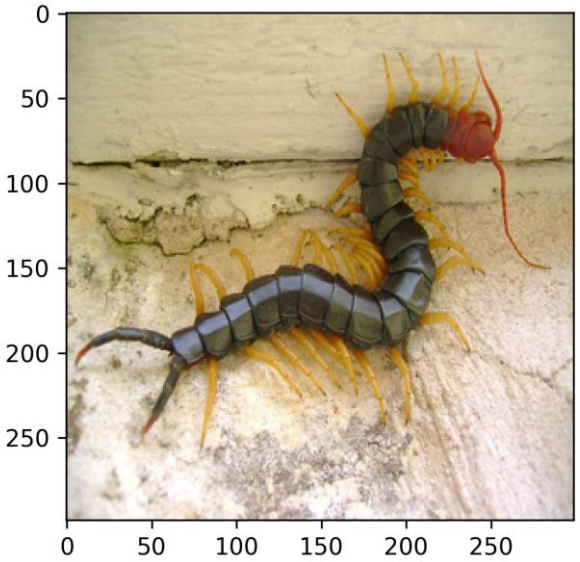
category: strawberry

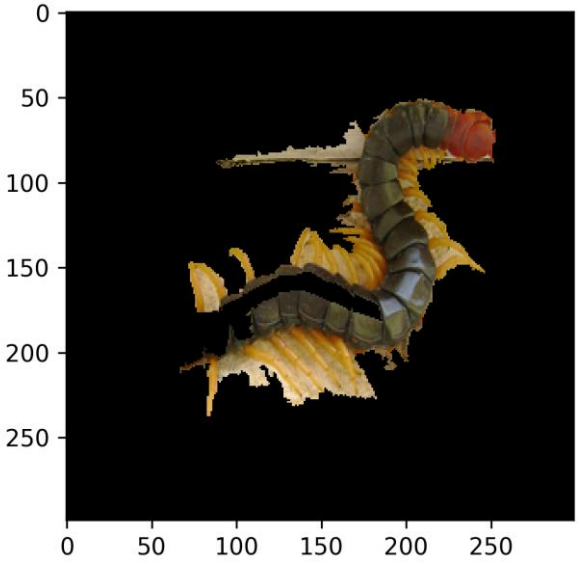

0
50
100
150
200
250
0
50
100
150
200
250

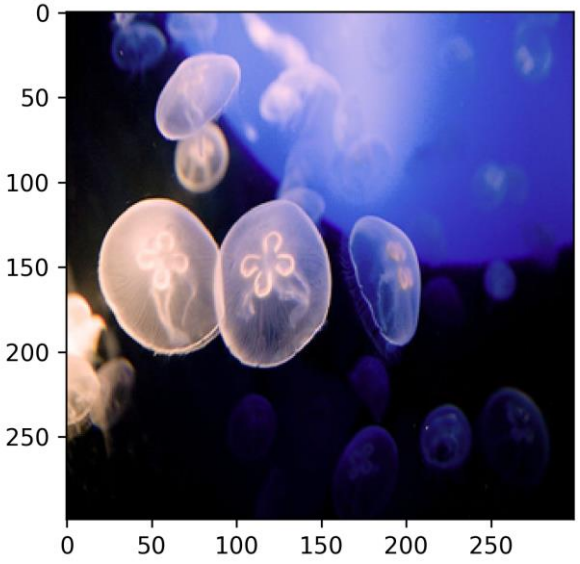

0
50
100
150
200
250
0
50
100
150
200
250

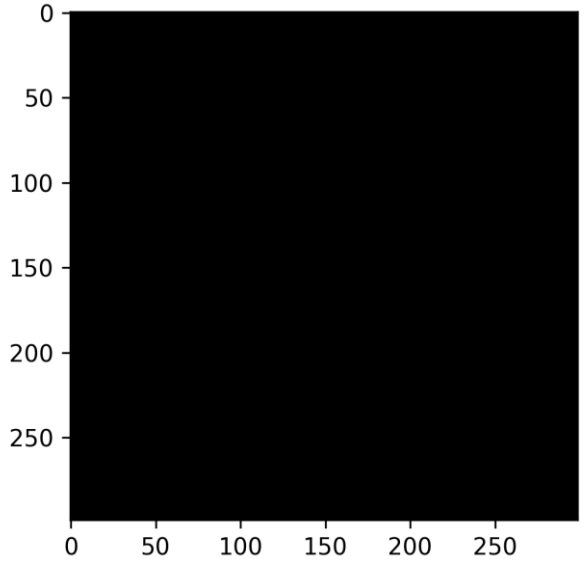
Download image from ImageNet

category: centipede

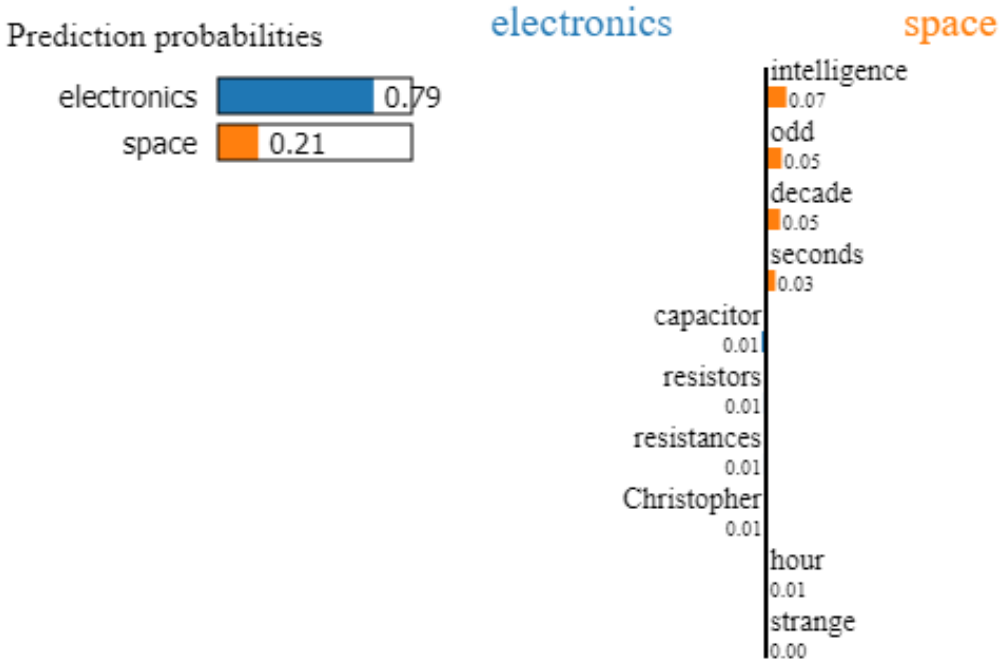



0
50
100
150
200
250
0
50
100
150
200
250


0
50
100
150
200
250
0
50
100
150
200
250


0
50
100
150
200
250
0
50
100
150
200
250

Chapter 12: Cognitive XAI



Edit - Datapoints 90 and 68

Search features

Feature	Value(s)	Counterfactual value(s)
image/encoded		
5_o_Clock_Shadow	0	0

Heavy_Makeup

High_Cheekbones

Male

Mouth_Slightly_ Open

Mustache

(none)

Datapoint ID

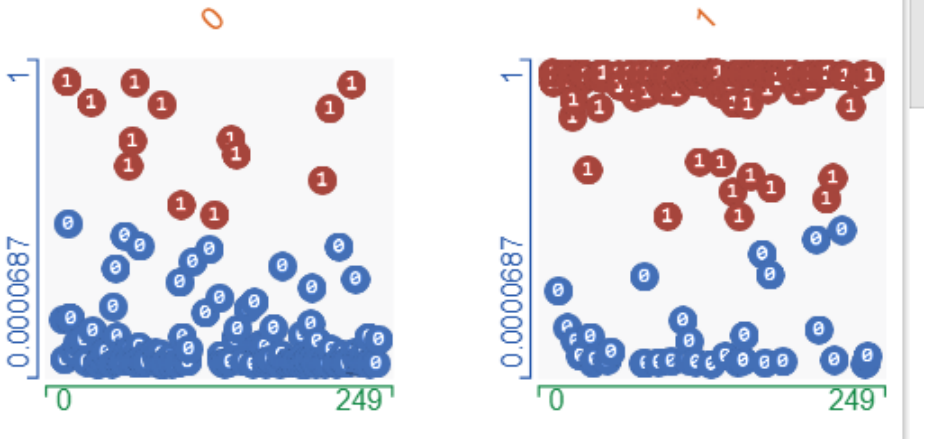
Inference score

(image)

Datapoint ID

Inference score

Inference value



Feature	Value(s)
image/encoded	