

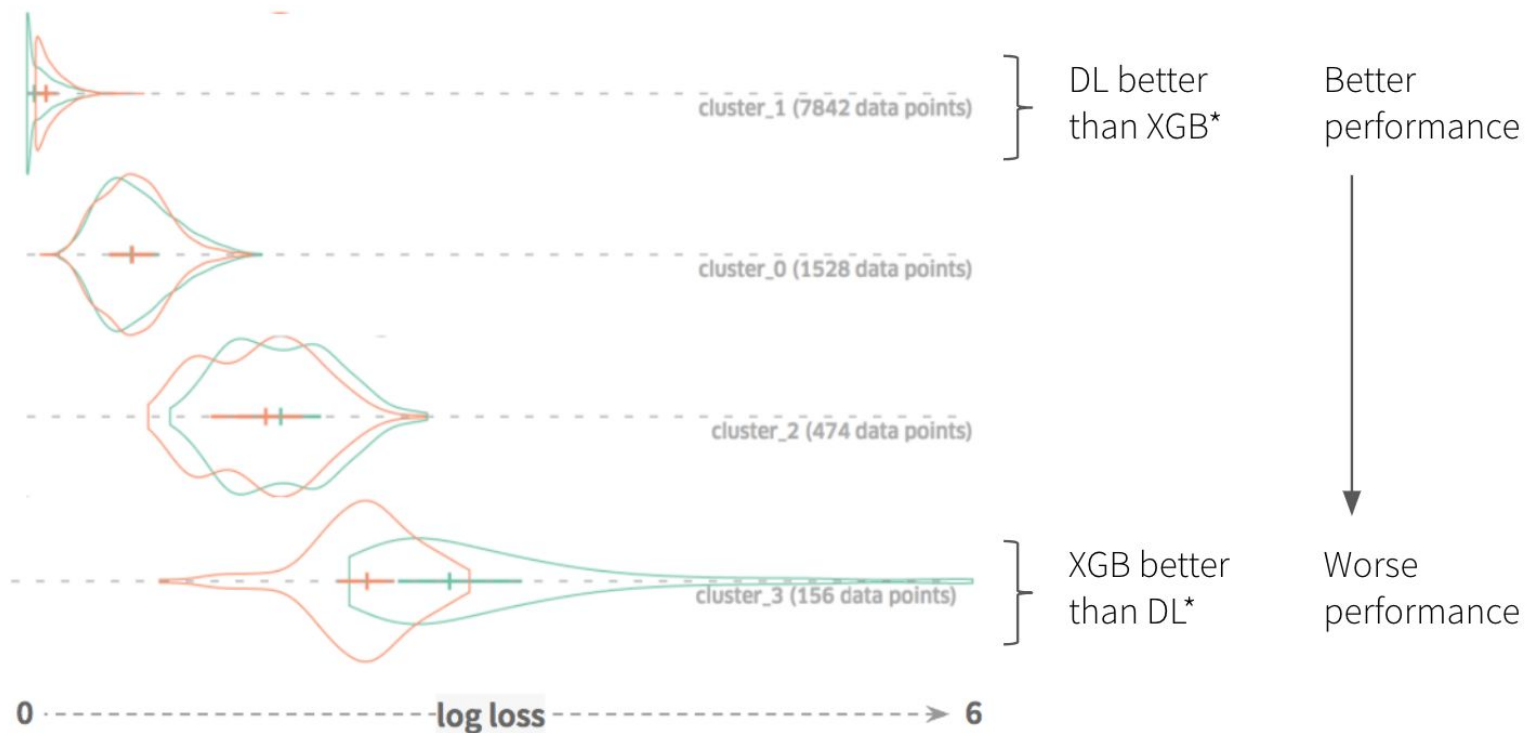
# Wizualna eksploracja modeli

Agnieszka Ciepielewska

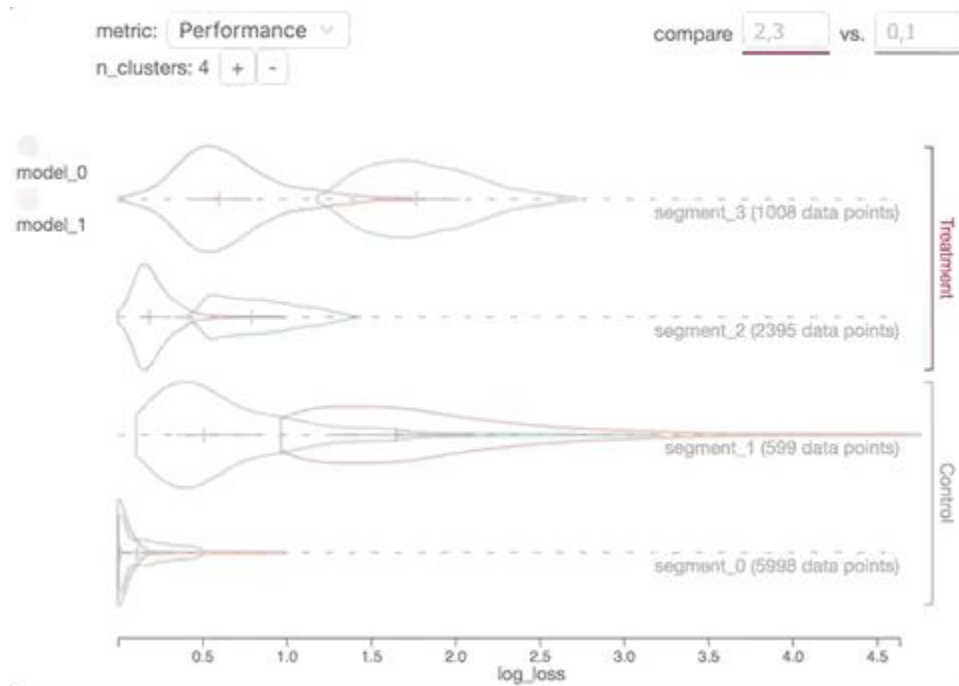
# Manifold

# Performance Comparison View

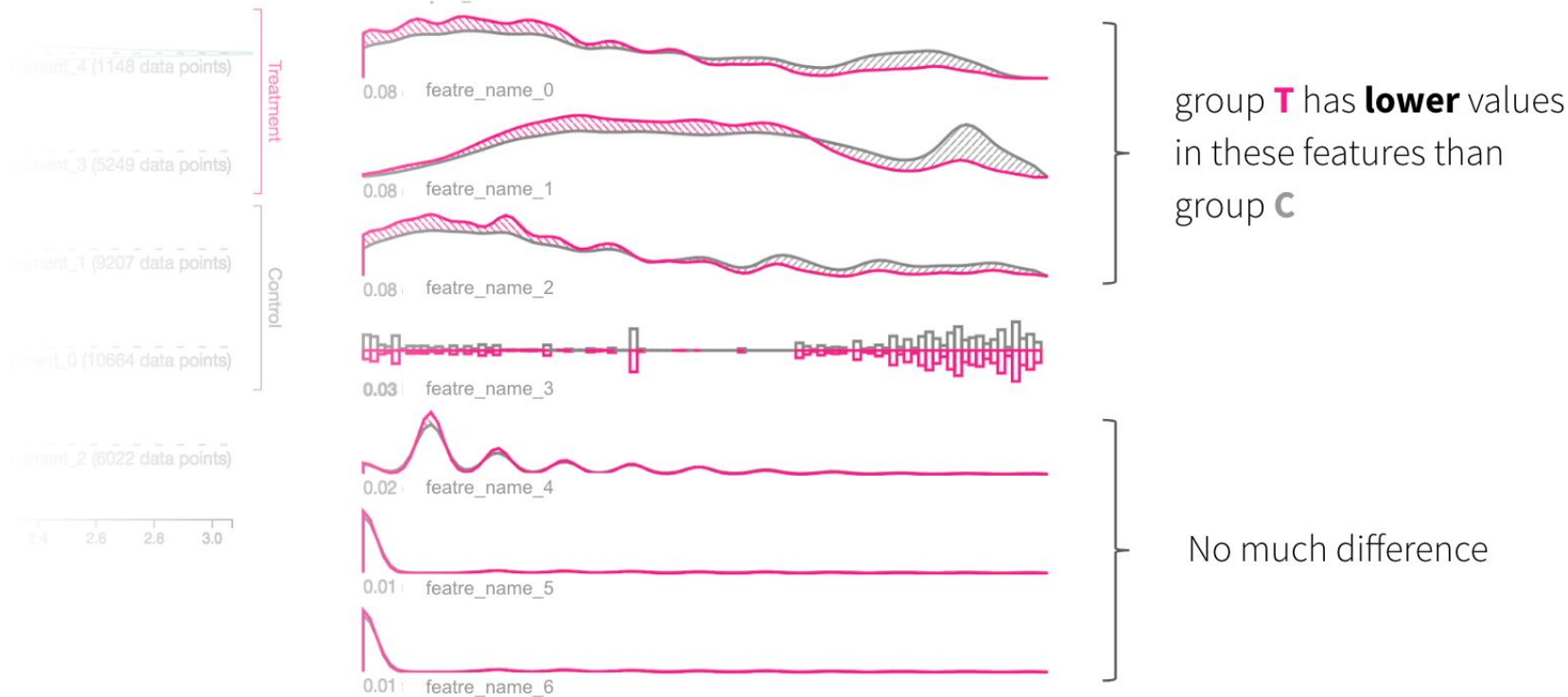
● DeepLearningV1  
● XGBoostV1



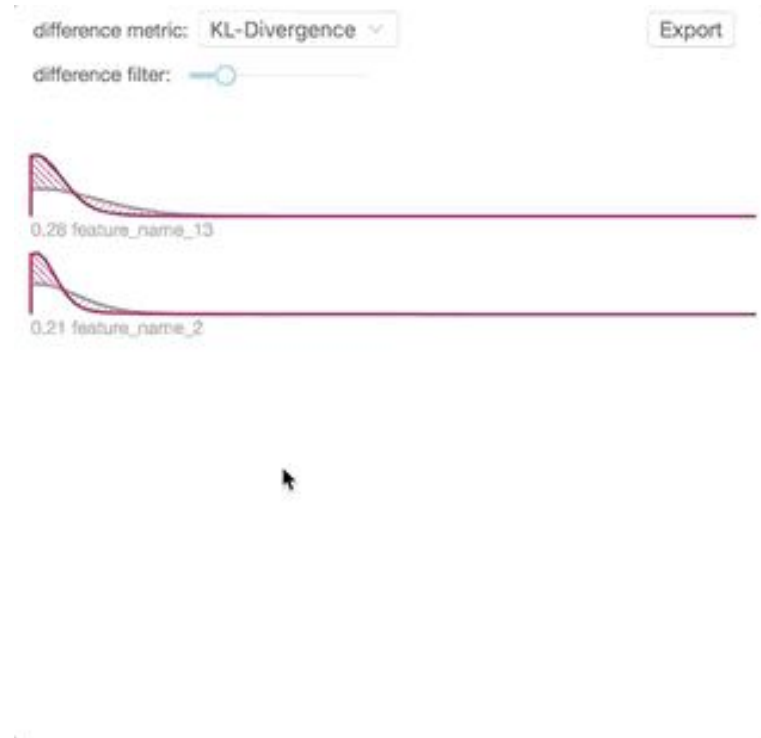
# Performance Comparison View



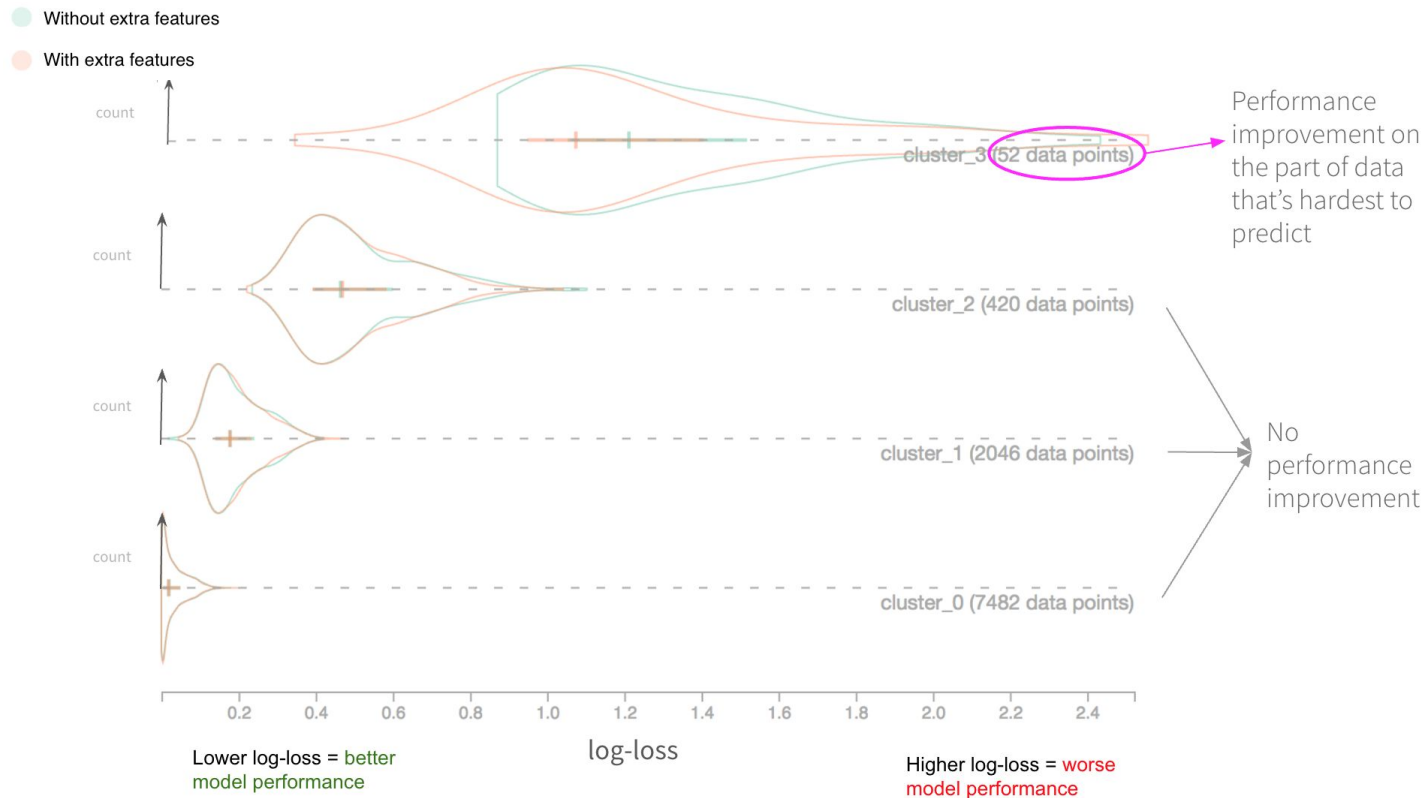
# Feature Attribution View



# Feature Attribution View

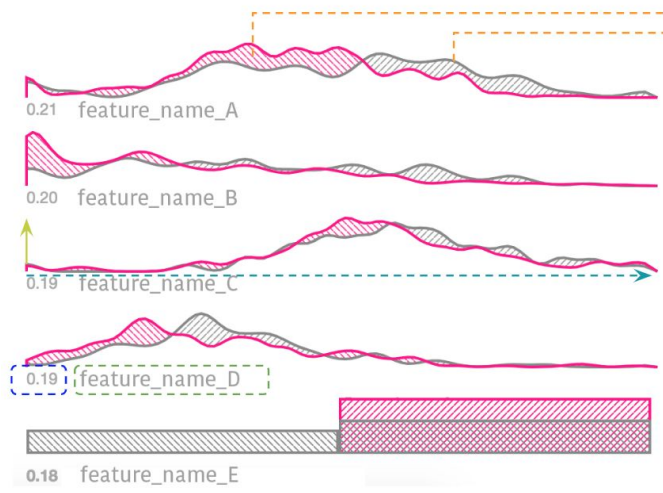
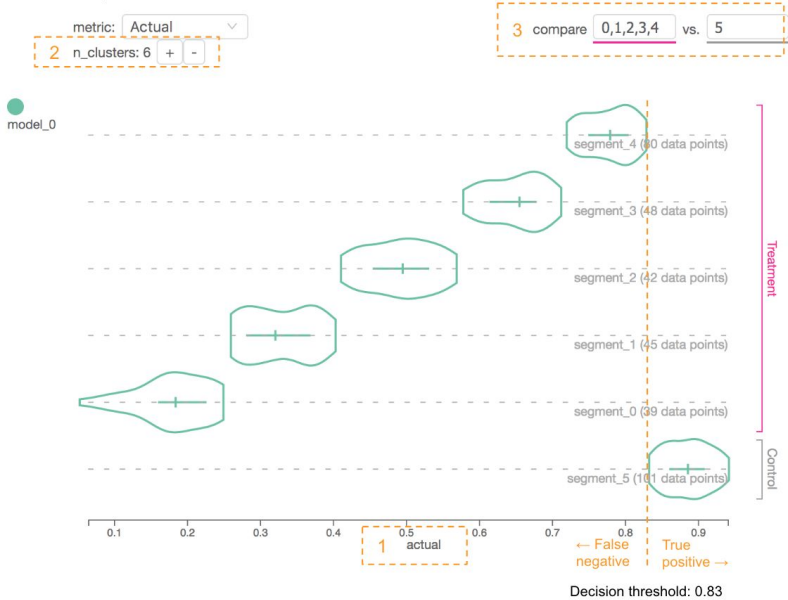


# Example of usage



# Example of usage

Dataset: all positive instances

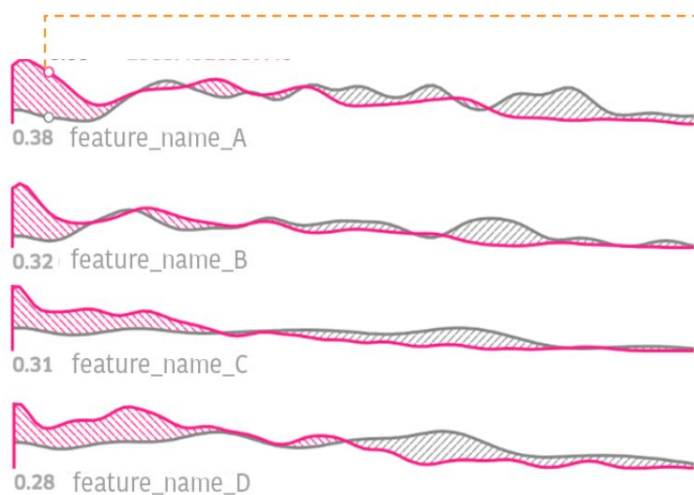


Distributions of these features show notable differences between the True Positive Group and the False Negative Group

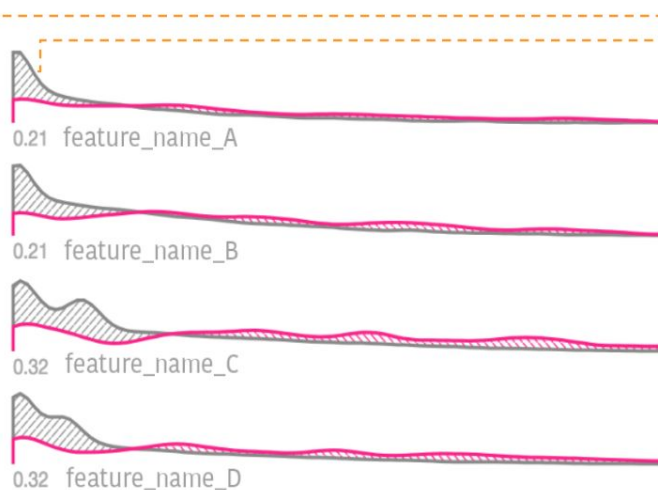
- KL-Divergence between 2 groups
- Feature names
- Feature value
- Number / percentage of data points



# Example of usage



Dataset: All **Positive** instances  
Compare **True Positive Group** and  
**False Negative Group**



Dataset: **All** instances  
Compare the **Negative Group** and  
the **Positive Group**

➤ Most negative instances tend to have low values in these features. Therefore, if an instance has low values in these features, the model tends to predict them as negative (which is sometimes wrong!)

## Plusy

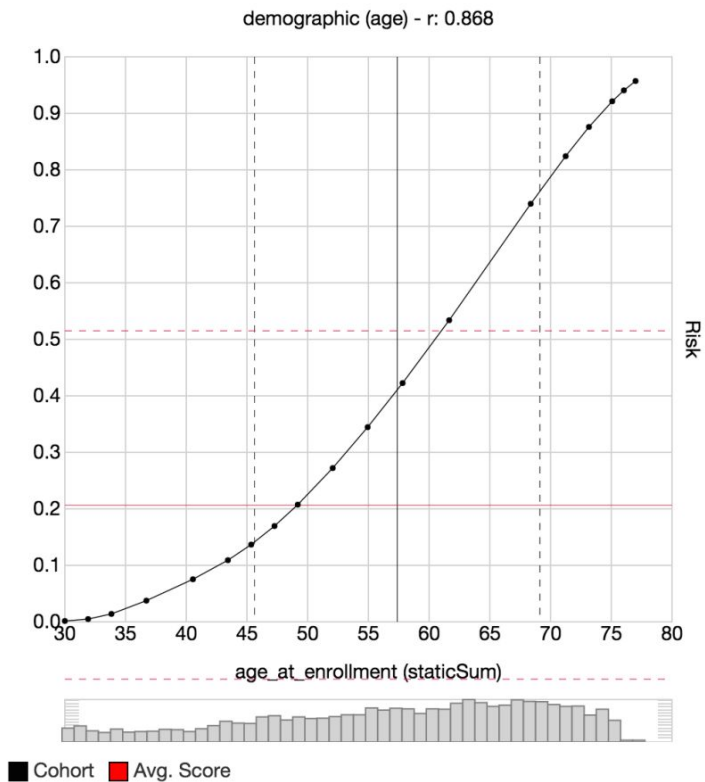
- Łatwe porównywanie modeli, które różnią się tylko dla trudnych obserwacji
- Ładny interface

## Minusy

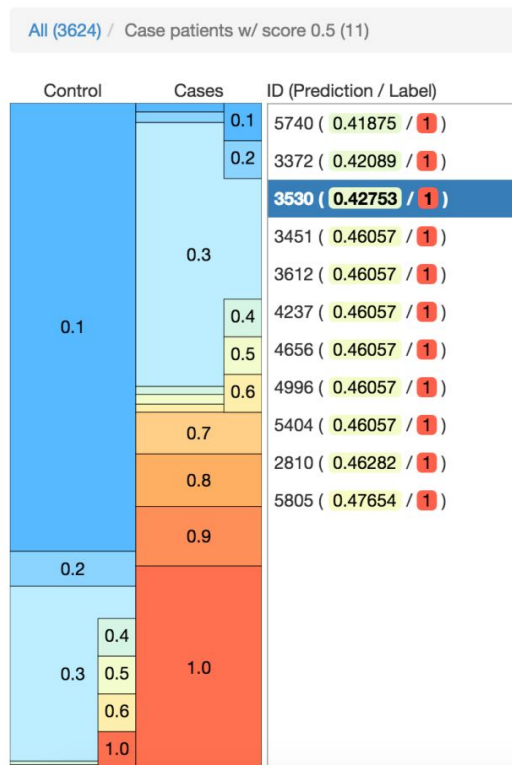
- Brak dostępnej implementacji
- Trudność porównywania modeli, które różnią się dla łatwych obserwacji

Prospector

# Partial dependence



# The interface for selecting an observation



# The summary of one observation

Patient: 3530 Truth: 1 Original: 0.42753

## Decreasing Risk:

Feature	Current	Suggested Change
<b>bmi (count)</b> vital (bmi)	0	1 ( 0.08021 )
<b>eGFR</b> lab	59.18887	59.59549 ( 0.23110 )
<b>bmi</b> vital (bmi)	28.27873	28.23937 ( 0.27954 )
<b>eGFR (count)</b> lab	0	1 ( 0.28705 )
<b>Calcifediol (Vit D) (25-O...</b>	0	1 ( 0.31857 )

## Increasing Risk:

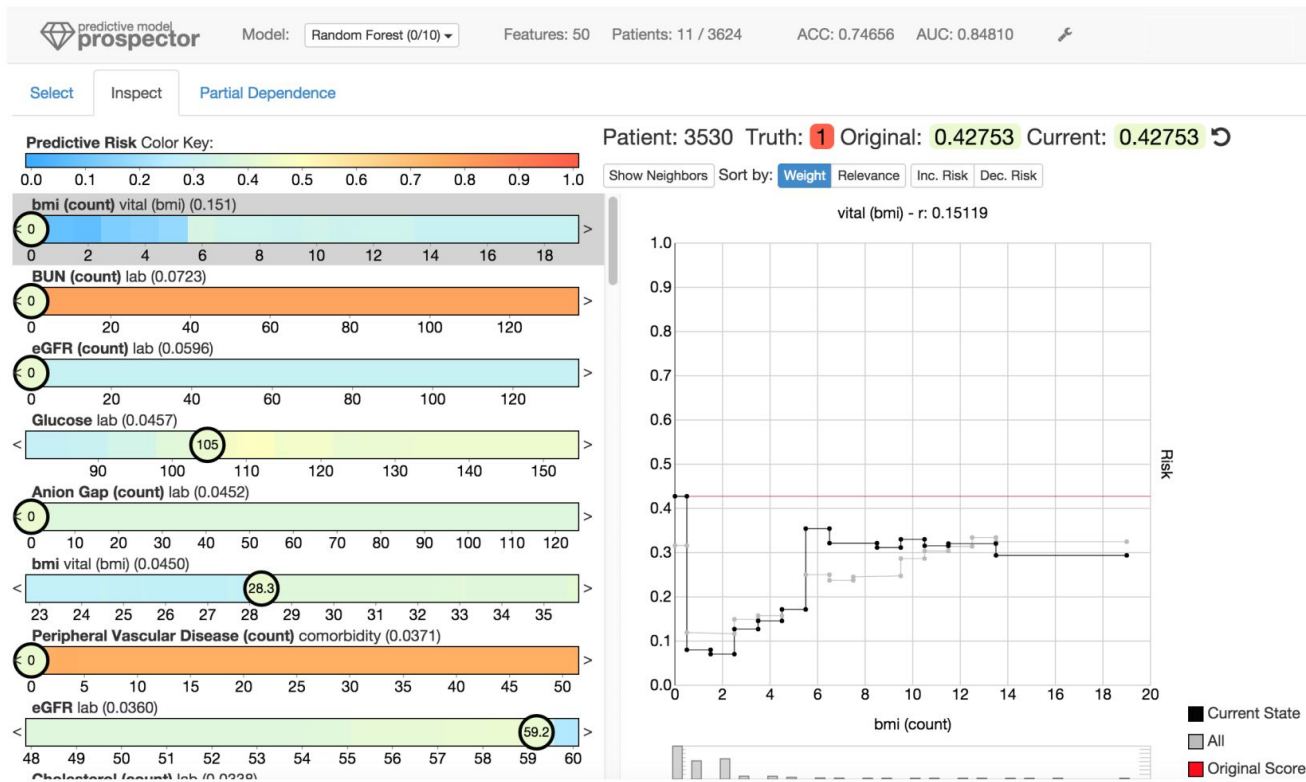
Feature	Current	Suggested Change
<b>BUN (count)</b> lab	0	1 ( 0.77246 )
<b>Peripheral Vascular Dis...</b>	0	1 ( 0.68666 )
<b>Uric Acid (count)</b> lab	0	1 ( 0.64202 )
<b>Calcium</b> lab	9.37486	9.38749 ( 0.59175 )
<b>Carbon Dioxide</b> lab	26.56109	27.35469 ( 0.59025 )

$$I_f(p) = \int_{-\infty}^{\infty} [pred(p^*) - pred(p)] \omega(v, p_f) dv$$

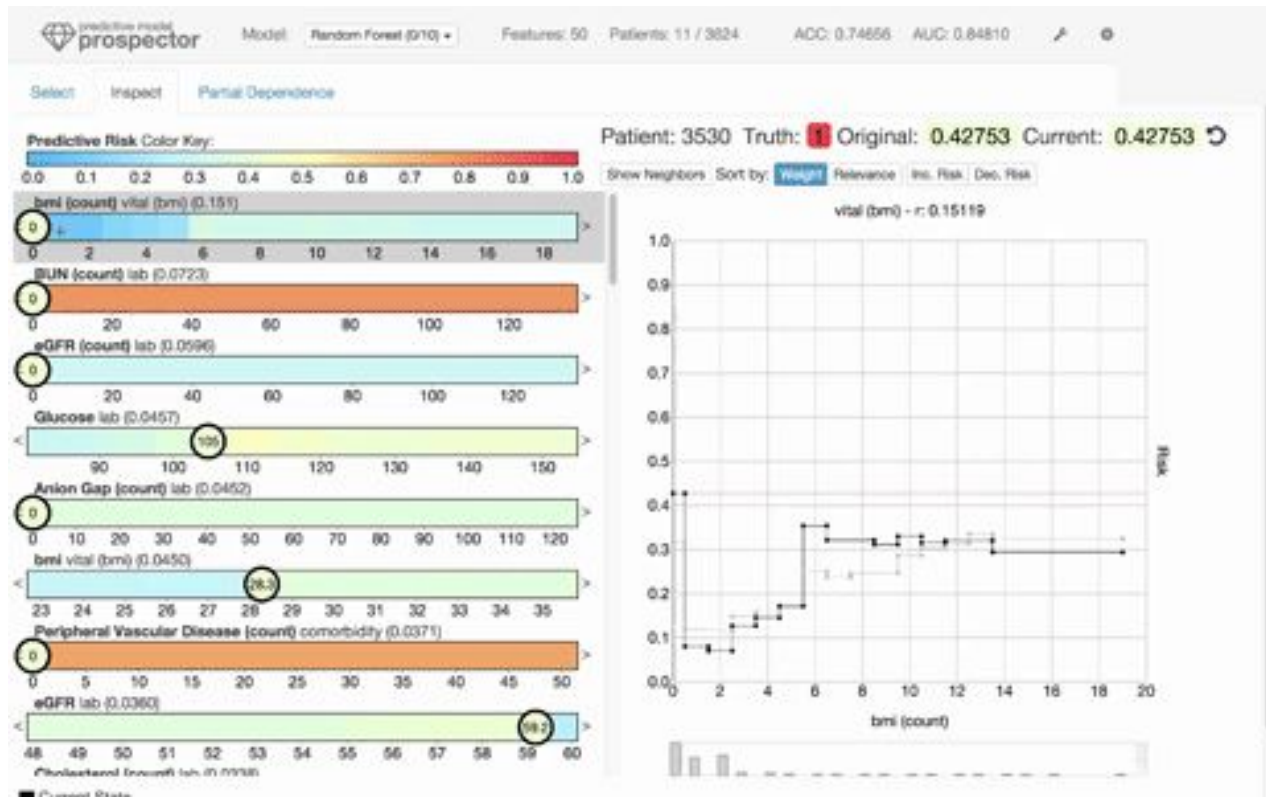
with  $p_f^* = v$  and  $p_g^* = p_g$  for  $g \neq f$

$$\omega(v, p_f) = \frac{1}{\sigma_f \sqrt{2\pi}} \exp\left(-\frac{(v - p_f)^2}{2\sigma_f^2}\right)$$

# Inspecting one observation



# Inspecting one observation





# Plusy

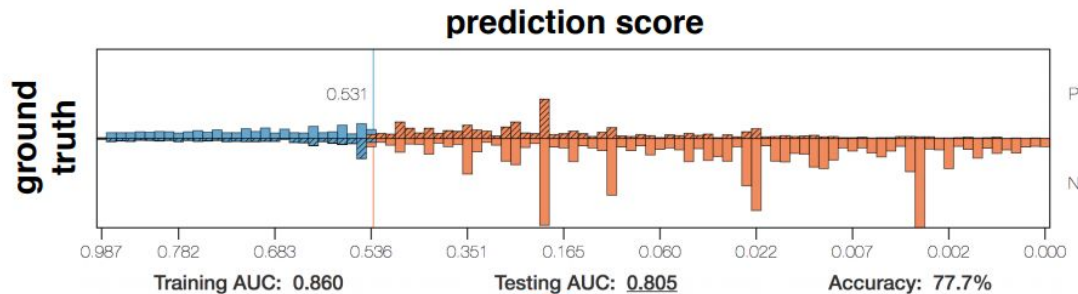
- Ładny interface do patrzenia na zmianę odpowiedzi modelu przy zmienianym inpucie

# Minusy

- Brak dostępnej implementacji
- Interface do inspekcji obserwacji wygląda jakby nie działał płynnie
- Nieintuicyjny interface do wybierania pojedynczej obserwacji

# Explanation explorer

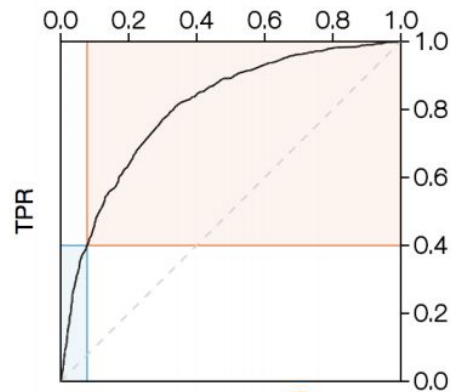
# Statistical Summary View



(A)

	P (prediction)	N (prediction)	
P (ground truth)	531	797	1328
N (ground truth)	269	3187	3456
	800	3984	4784

(B)



FPR (C)

## Controls & Filters

Q A

4784
<span style="border: 1px solid black; border-radius: 50%; padding: 2px 5px;">B</span> 2201
+

Filter: C

☐ total > 20

Order: D

☒ total

☐ positive

☐ negative

☐ true positive

☐ false negative

☐ false positive

☐ true negative

☐ incorrect

☐ correct

☐ explanation length

☐ explanation

## Explanations

## Statistics about explained items

		tp	fp	fn	tn	pos / total	odds ratio	
		1419				624 / 2201	0.1 1 10	
	SODIUM CHLORIDE	93		207		99 / 313		=>
	IBUPROFEN		218			13 / 231		=>
	VANCOMYCIN	136		49		136 / 185		=>
	ACETAMINOPHEN		137			22 / 159		=>
	ASPIRIN	63		68		63 / 131		=>
	OXYCODONE-ACET.		102			14 / 116		=>
	SODIUM CHLORIDE ONDANSETRON		101			14 / 115		=>
	DIA. MEGLUMINE.	19	26		46	31 / 103		=>
	KETOROLAC		84			13 / 97		=>
	MORPHINE	29		45		29 / 74		=>
	SODIUM CHLORIDE ACETAMINOPHEN	13		40		13 / 53		=>
	SODIUM CHLORIDE DIA. MEGLUMINE.	19		27		19 / 46		=>
	<span style="border: 1px solid black; border-radius: 50%; padding: 2px 5px;">E</span> FAMOTIDINE <span style="border: 1px solid black; border-radius: 50%; padding: 2px 5px;">F</span>	9		36		<span style="border: 1px solid black; border-radius: 50%; padding: 2px 5px;">G</span> 9 / 45 <span style="border: 1px solid black; border-radius: 50%; padding: 2px 5px;">H</span>		<span style="border: 1px solid black; border-radius: 50%; padding: 2px 5px;">I</span> =>
	FUROSEMIDE		36	8		36 / 44		=>
	ONDANSETRON MORPHINE DIA. MEGLUMINE. 1>	25		18		25 / 43		=>
	LIDOCAINE		40			3 / 43		=>
	ONDANSETRON	13		28		13 / 41		=>
	SODIUM CHLORIDE MORPHINE	15		24		15 / 39		=>
	SODIUM CHLORIDE ONDANSETRON MORPHINE	10		27		10 / 37		=>
	DIPH	3		31		3 / 34		=>
	SODIUM CHLORIDE ONDANSETRON FAMOTIDINE	4		27		4 / 31		=>
	SODIUM CHLORIDE KETOROLAC		27			2 / 29		=>
	SODIUM CHLORIDE ONDANSETRON DIA. MEGLUMINE.	10		15		10 / 25		=>
	ALBUTE. SULFATE	10		15		10 / 25		=>
	SODIUM CHLORIDE IPRATR. BROMIDE ALBUTE. SULFATE	7		18		7 / 25		=>
	METOCLOPRAMIDE KETOROLAC		25			0 / 25		=>
	OXYCODONE-ACET. IBUPROFEN		25			0 / 25		=>
	DIAZEPAM	3		20		3 / 23		=>
	ONDANSETRON MORPHINE KETOROLAC 1>		23			0 / 23		=>
	ONDANSETRON MORPHINE	8		13		8 / 21		=>

Features needed to be removed  
to change the predicted label of items

Distribution  
of Items

Number  
of Items

Odds Ratio  
on Log Scale



## Controls & Filters

Q A v

4784
<span style="border: 1px solid black; border-radius: 50%; padding: 2px;">B</span> 2201
+

Filter: C

☐ total > < 20

Order: D

- ☐ total
- ☐ positive
- ☐ negative
- ☐ true positive
- ☐ false negative
- ☐ false positive
- ☐ true negative
- ☐ incorrect
- ☐ correct
- ☐ explanation length
- ☐ explanation

## Explanations

## Statistics about explained items

		tp	fp	fn	tn	pos / total	odds ratio	
		1419				624 / 2201	0.1 1 10	
SODIUM CHLORIDE	SODIUM CHLORIDE	93		207		99 / 313		=>
	IBUPROFEN		218			13 / 231		=>
	VANCOMYCIN	136		49		136 / 185		=>
	ACETAMINOPHEN	137				22 / 159		=>
	ASPIRIN	63		68		63 / 131		=>
	OXYCODONE-ACET.		102			14 / 116		=>
	ONDANSETRON		101			14 / 115		=>
	DIA. MEGLUMINE.	19	26	46		31 / 103		=>
	KETOROLAC		84			13 / 97		=>
	MORPHINE	29		45		29 / 74		=>
SODIUM CHLORIDE	ACETAMINOPHEN	13		40		13 / 53		=>
SODIUM CHLORIDE	DIA. MEGLUMINE.	19		27		19 / 46		=>
ONDANSETRON	FAMOTIDINE	9		36		9 / 45		=>
	FUROSEMIDE	36		8		36 / 44		=>
	DIA. MEGLUMINE. 1>	25		18		25 / 43		=>
	LIDOCAINE		40			3 / 43		=>
	ONDANSETRON	13		28		13 / 41		=>
	SODIUM CHLORIDE	15		24		15 / 39		=>
	MORPHINE	10		27		10 / 37		=>
	DIPH	3		31		3 / 34		=>
	ONDANSETRON	4		27		4 / 31		=>
	SODIUM CHLORIDE	7		27		2 / 29		=>
SODIUM CHLORIDE	ONDANSETRON	10		15		10 / 25		=>
SODIUM CHLORIDE	ONDANSETRON	10		15		10 / 25		=>
SODIUM CHLORIDE	ALBUTE. SULFATE	10		15		10 / 25		=>
	IPRATR. BROMIDE	7		18		7 / 25		=>
	METOCLOPRAMIDE		25			0 / 25		=>
	KETOROLAC		25			0 / 25		=>
	OXYCODONE-ACET.		25			0 / 25		=>
	IBUPROFEN	3		20		3 / 23		=>
	DIAZEPAM		23			0 / 23		=>
	KETOROLAC		23			0 / 23		=>
	ONDANSETRON	8		13		8 / 21		=>
	MORPHINE							=>

Features needed to be removed  
to change the predicted label of items

Distribution  
of Items

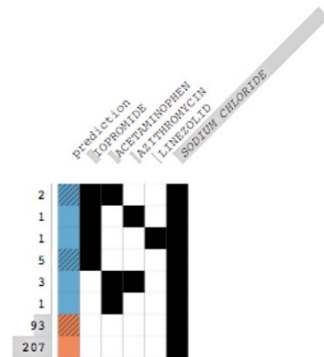
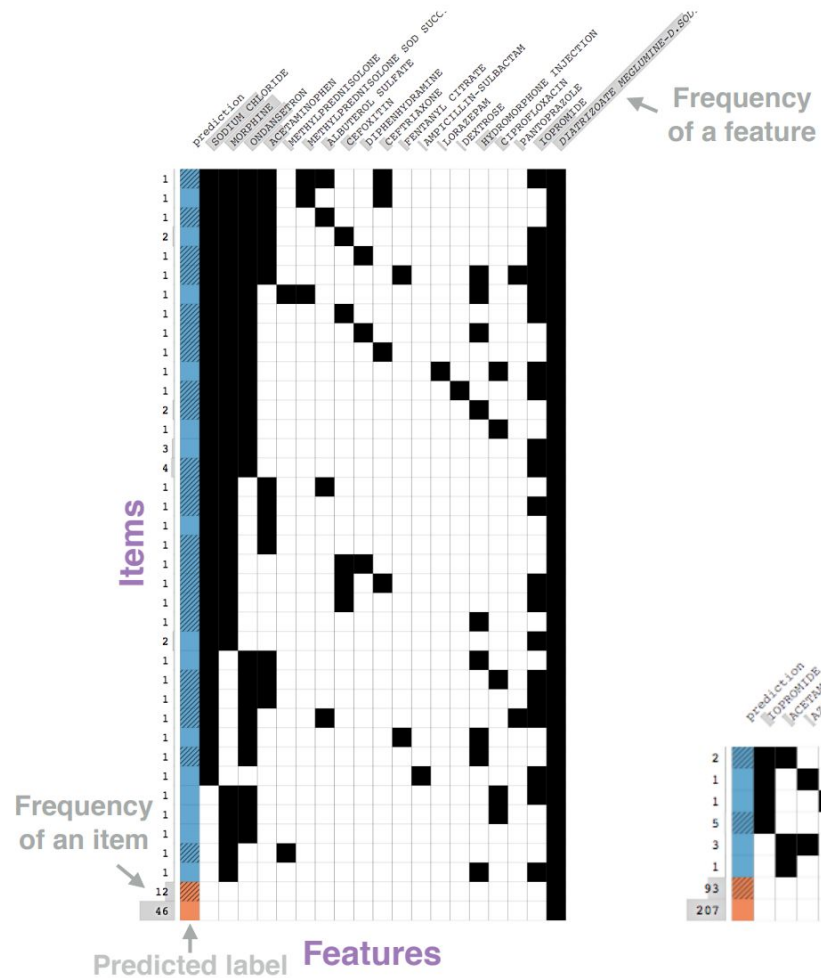
Number  
of Items

Odds Ratio  
on Log Scale

- true positive
- false positive
- false negative
- true negative
- positive
- negative

$$\min_e L(v - e) \neq L(v)$$

$$\frac{p_e/n_e}{p_t/n_t}$$



# Plusy

- Dostępna implementacja na githubie

# Minusy

- Implementacja z githuba nie działa
- Działa tylko dla binarnej klasyfikacji
- Mało intuicyjne tworzenie raportu
- Bardzo długi czas generowania raportu

Model describer



# Impact

## Model-Describer - Impact By Variable

[How to interpret Model-Describer Charts](#)

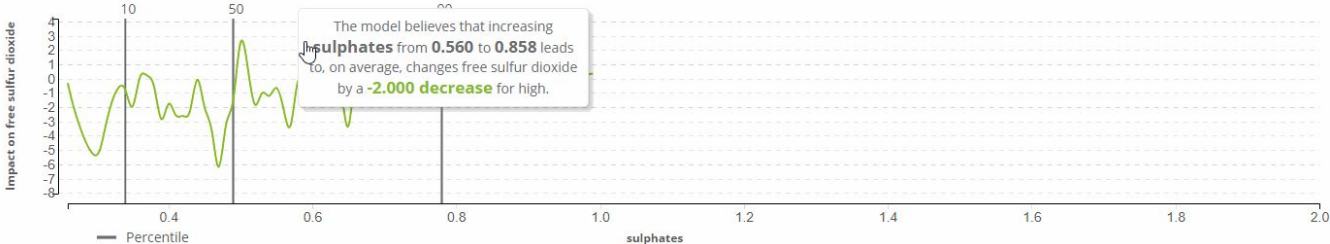
Category: low (50.716%) Average Error: 0.030 Error Type: MEAN	Category: medium (43.805%) Average Error: 0.020 Error Type: MEAN
---	--

Group By: alcohol ▼

☒ Percentiles

Impact of increasing sulphates by 0.298 on free sulfur dioxide

high



- Across all categories, the impact of **sulphates** on **free sulfur dioxide** is lowest from [0.39,0.49] for high and highest from [0.6,0.99] for high
- The impact of **sulphates** on **free sulfur dioxide** from [0.6,0.99] for high is 0.919 higher than the overall median impact and 3.287 higher than the average impact from [0.39,0.49] for high

# Error

## Model-Describer - Prediction Error By Variable

[How to interpret model-describer Charts](#)

Category: Red (24.611%)  
Average Error: 0.070

Category: White (75.389%)  
Average Error: 0.070

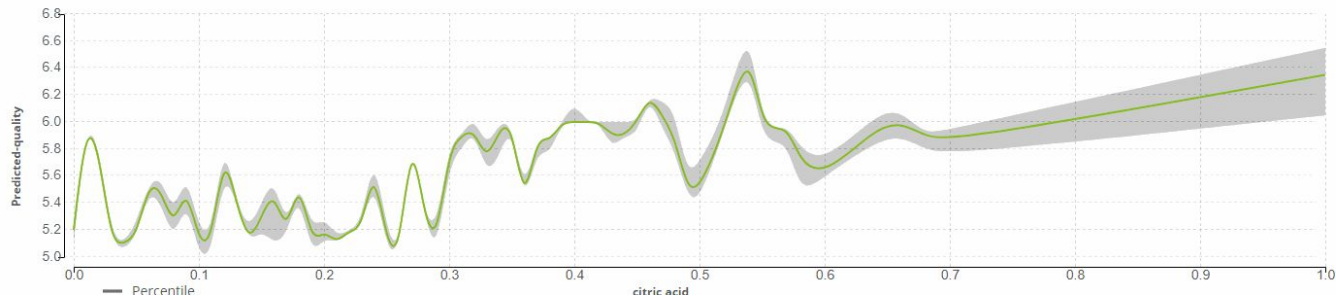
Group By: Type

Percentiles

Group **White** is 75.389% of the data and has an average error of **0.070**

Error Distribution Graph: citric acid (Grouped by Type)

Red



- Across all categories, the model's error in predicting **quality** is lowest from **[0.09,0.26]** for Red and highest from **[0.42,1]** for Red
- The error from **[0.42,1]** for Red is 0.237 higher than the overall median error and 0.649 higher than the average error from **[0.09,0.26]** for Red
- When the model makes prediction errors, there are times when those errors systematically lean in one direction
- The model tends to consistently make mistakes in the same direction when it miss-estimates citric acid. Within the Red group, the model consistently makes overestimates in the quartile range(s) **0.000-0.090, 0.090-0.260, 0.260-0.420, 0.420-1.000**

# Plusy

- Dostępna implementacja na githubie
- Działa, wystarczy zainstalować pakiet z githuba
- Działa dla różnych rodzajów modeli (regresja, klasyfikacja)
- Generuje mały plik html, który nie ma żadnych zależności
- Generuje się szybko

# Minusy

- Brak innowacyjnych rozwiązań
- Przesadne wygładzenie
- Niezrozumiałe opisy słowne
- Brak wykresów błędów (same wykresy predykcji)

# Linki do omówionych platform

1. Manifold: <https://arxiv.org/pdf/1808.00196.pdf>
2. Prospector: <https://josuakrause.github.io/info/material/chi2016-prospector.pdf>
3. Explanation explorer: [https://github.com/nyuvis/explanation\\_explorer](https://github.com/nyuvis/explanation_explorer)
4. Model describer: <https://github.com/DataScienceSquad/model-describer>