



# Тестирование продуктов с ML: Black box под капотом

---

Иван Ямщиков, PhD

ABBYY AI-evangelist

Софья Селезнева,

ABBYY QA-engineer



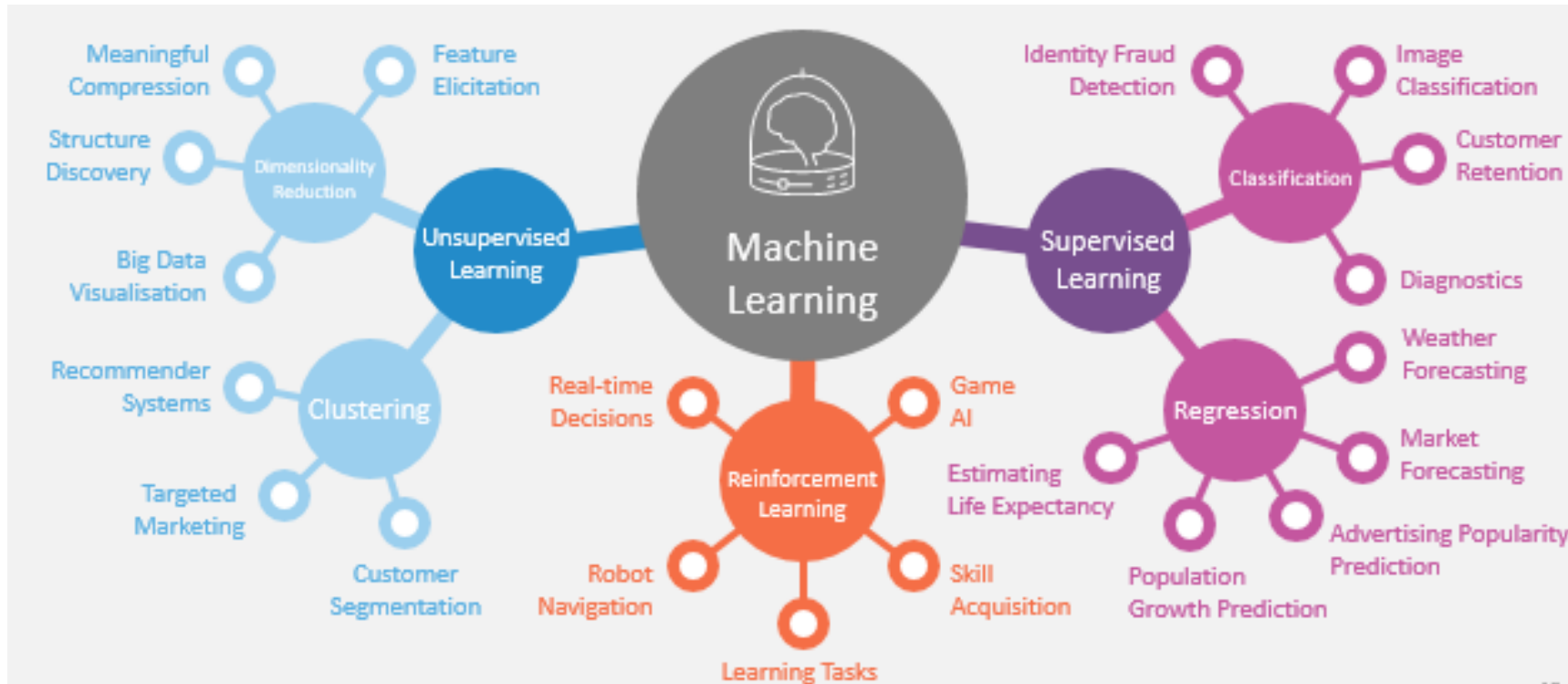
# О чём поговорим?

- Что такое ML?
- Как выглядит QA для специалиста по данным?
- Как выглядит QA здорового человека?
- Как наводить порядок в системе с ML внутри?

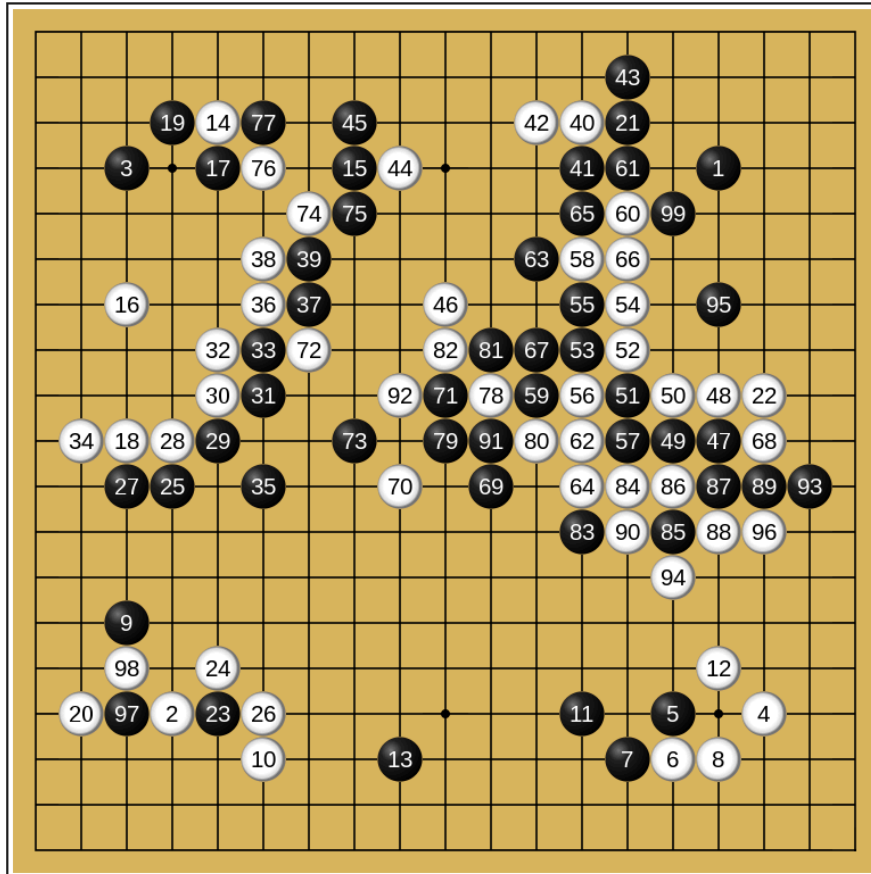
# Где мы применяем машинное обучение?



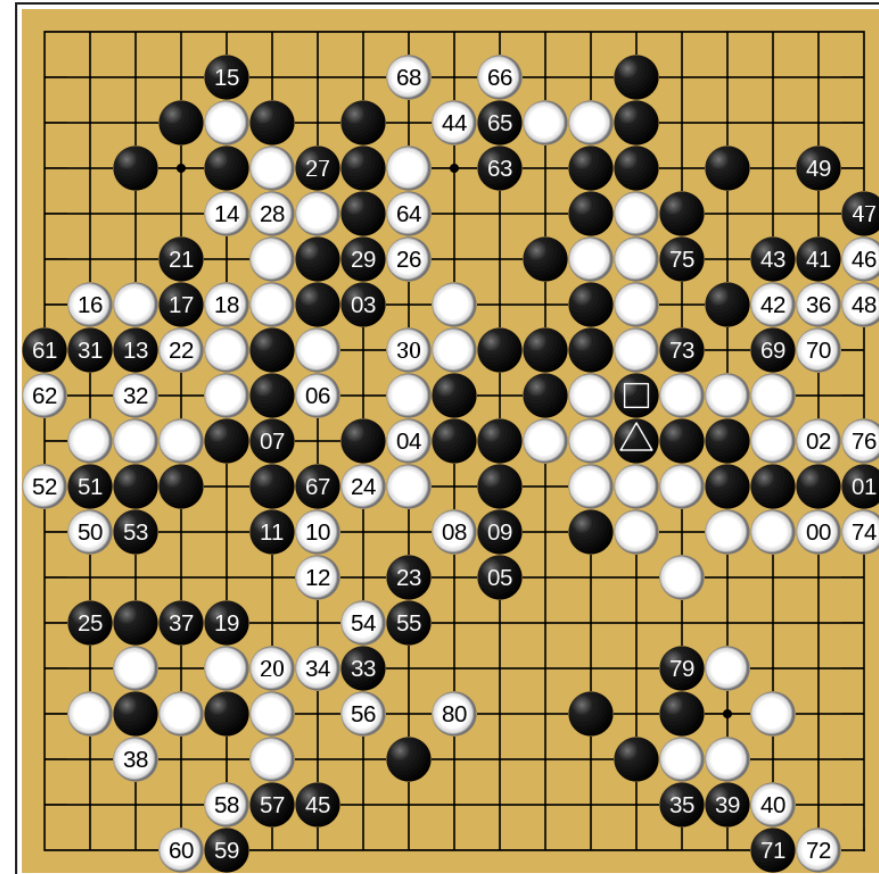
# Что за зверь такой? Машинное обучение



# Обучение с подкреплением

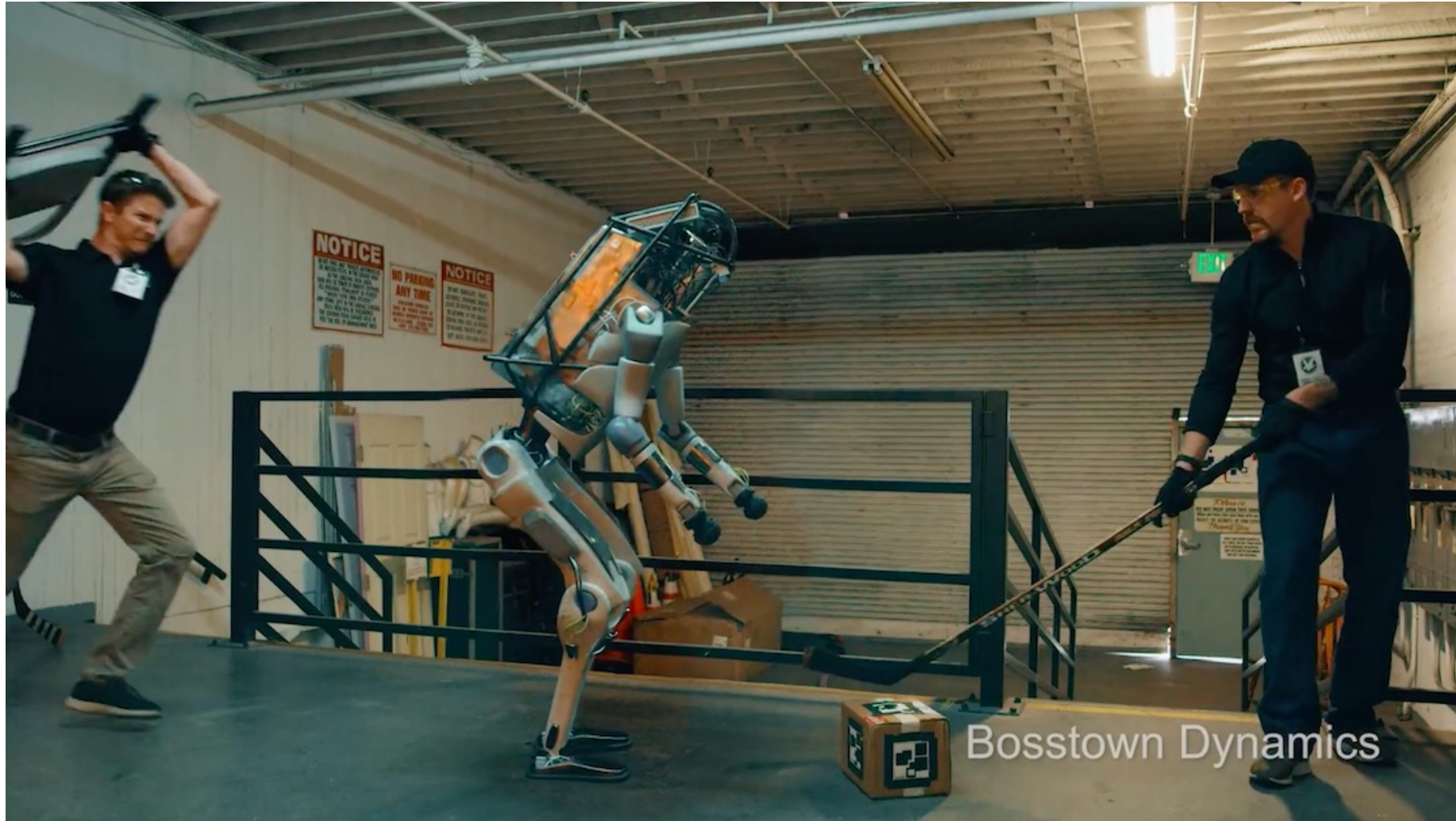


First 99 moves



Moves 100–180 (177 at , 178 at )

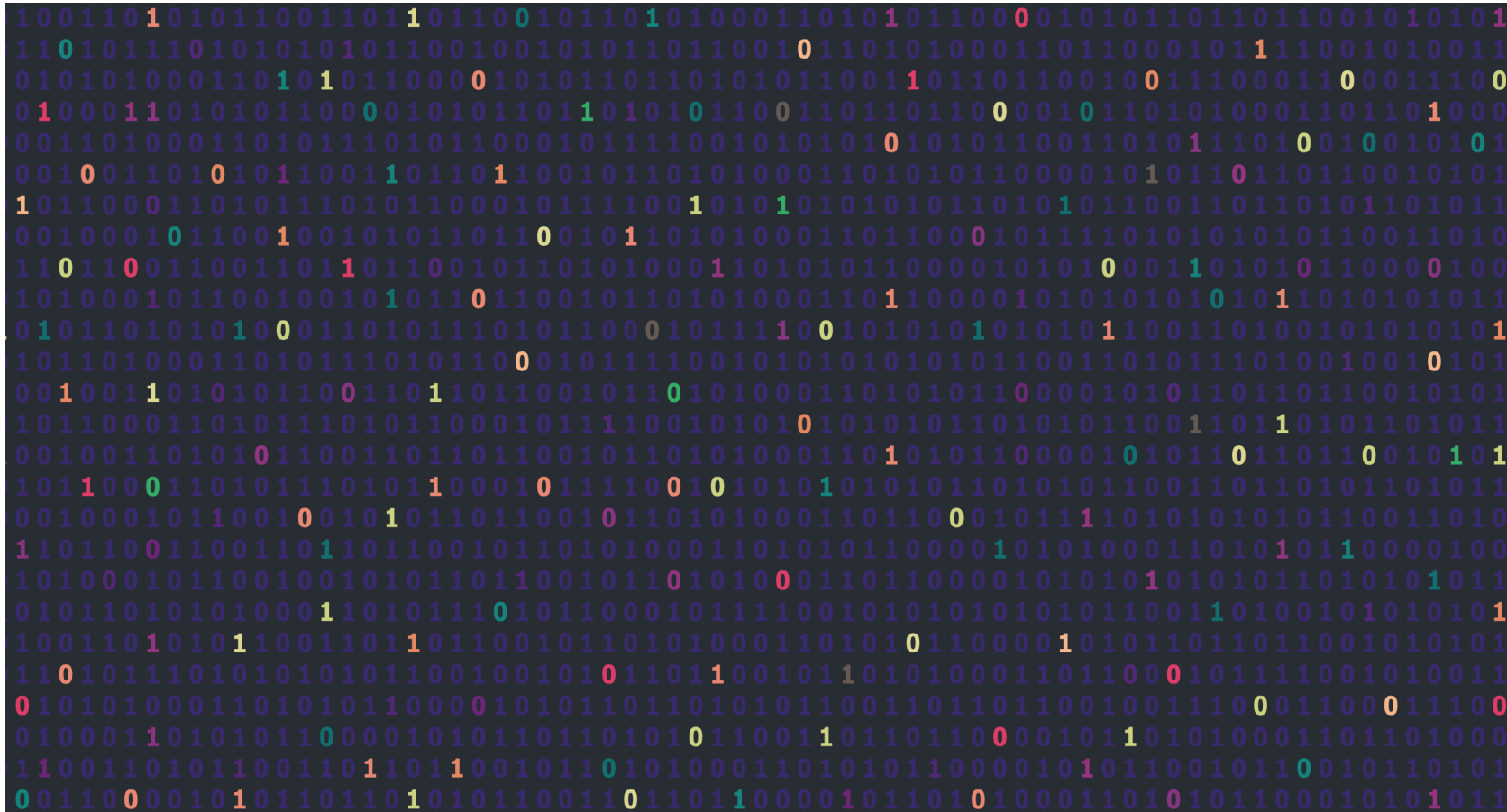
# Обучение с подкреплением



# Обучение с подкреплением



# Обучение без учителя

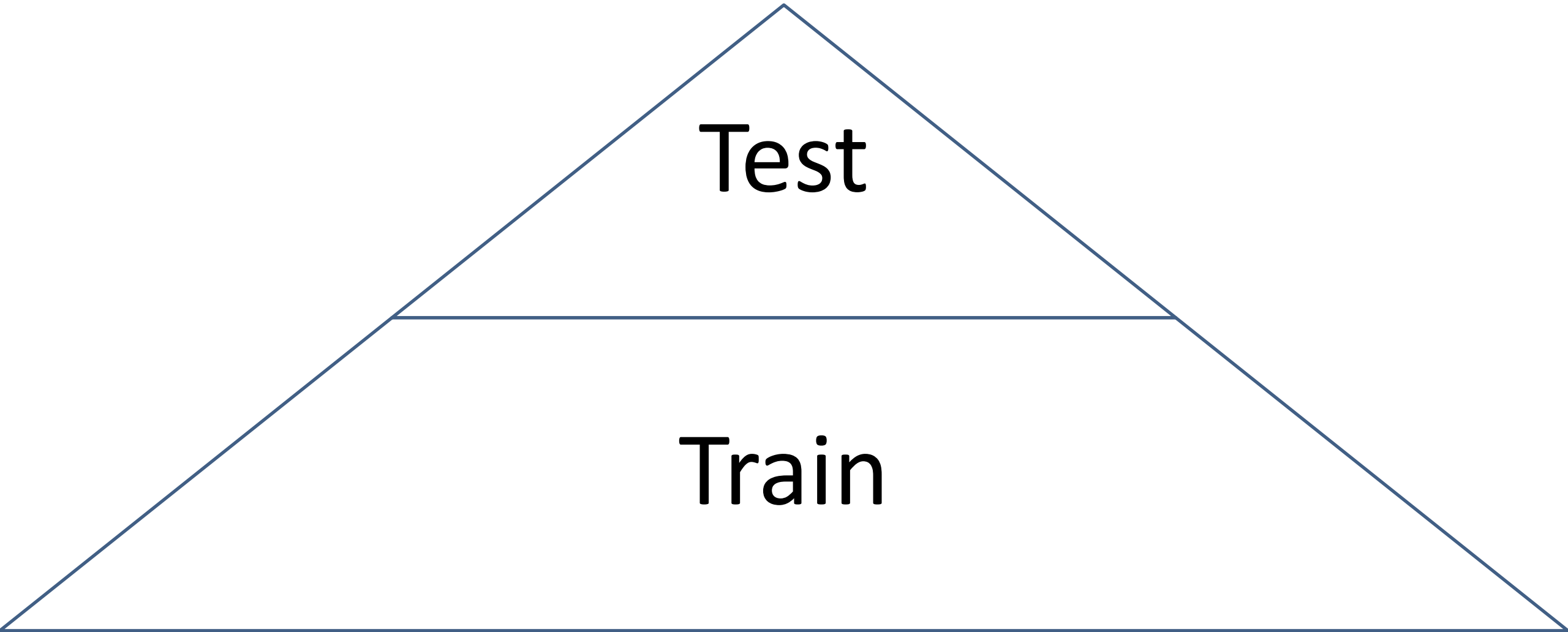




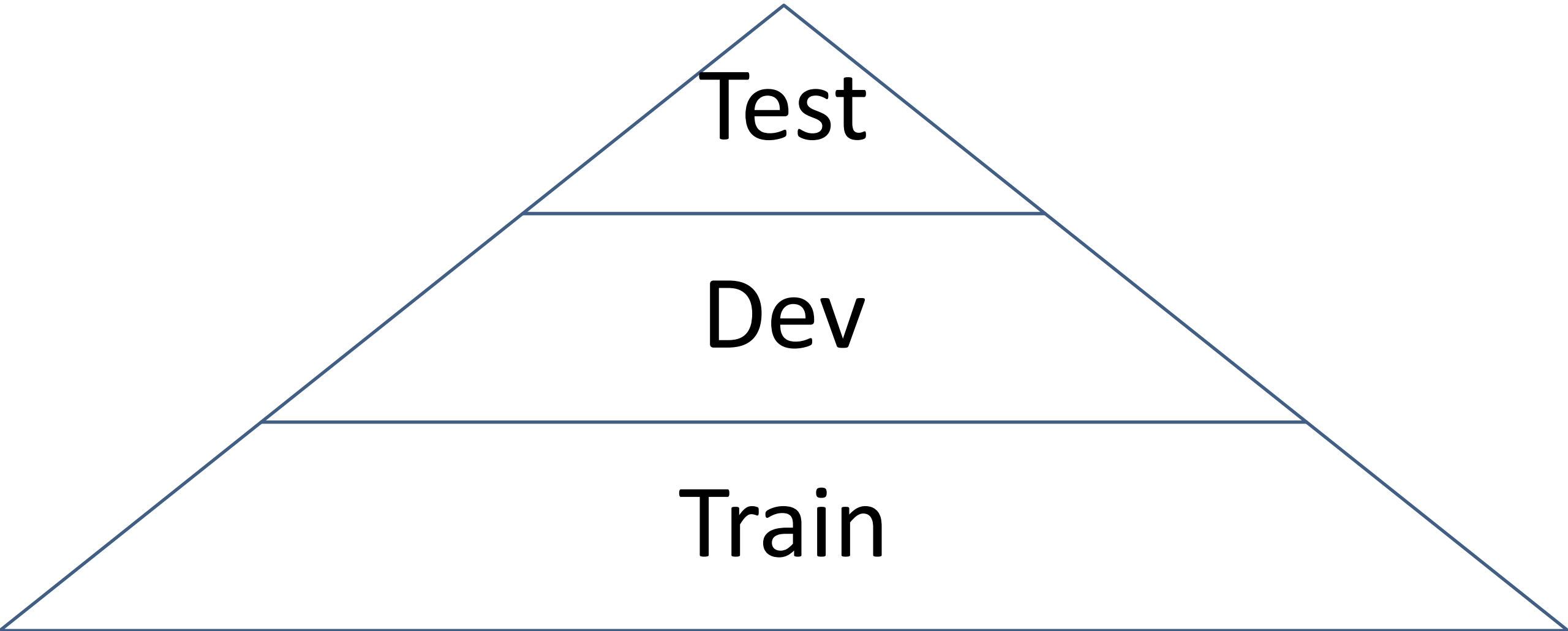
# Обучение с учителем



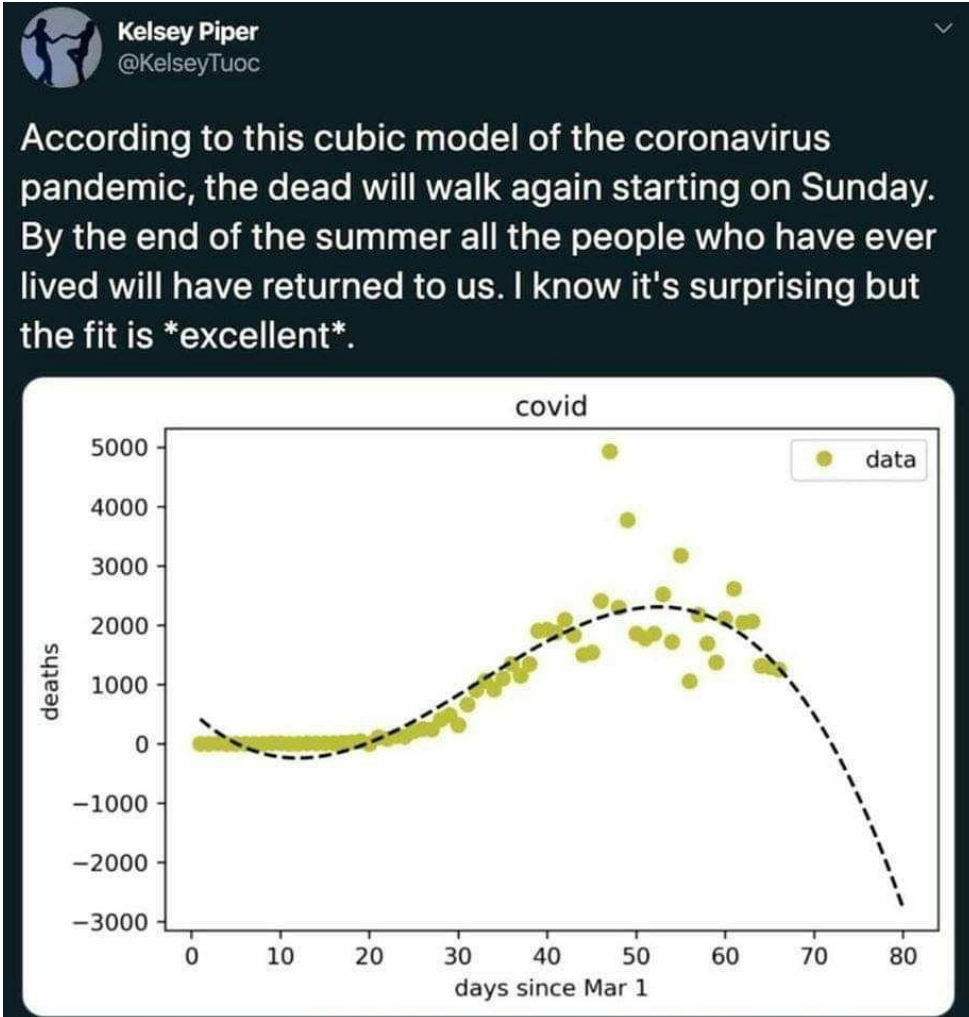
Как тестируют обучение с учителем?



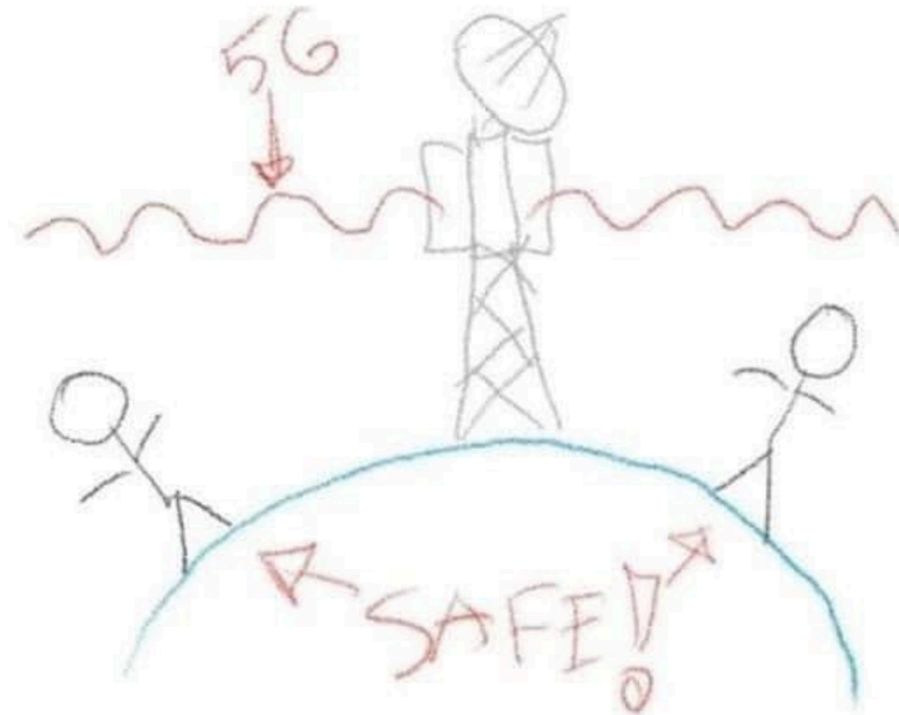
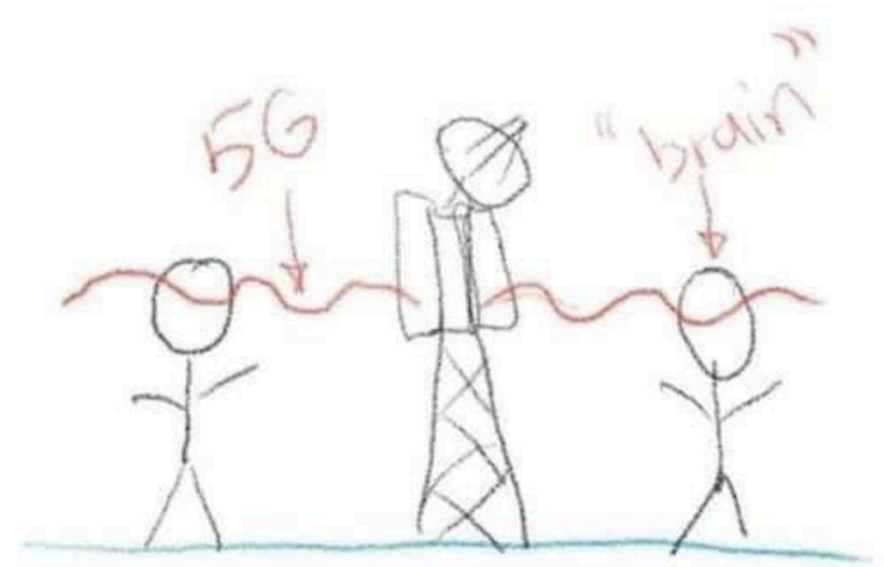
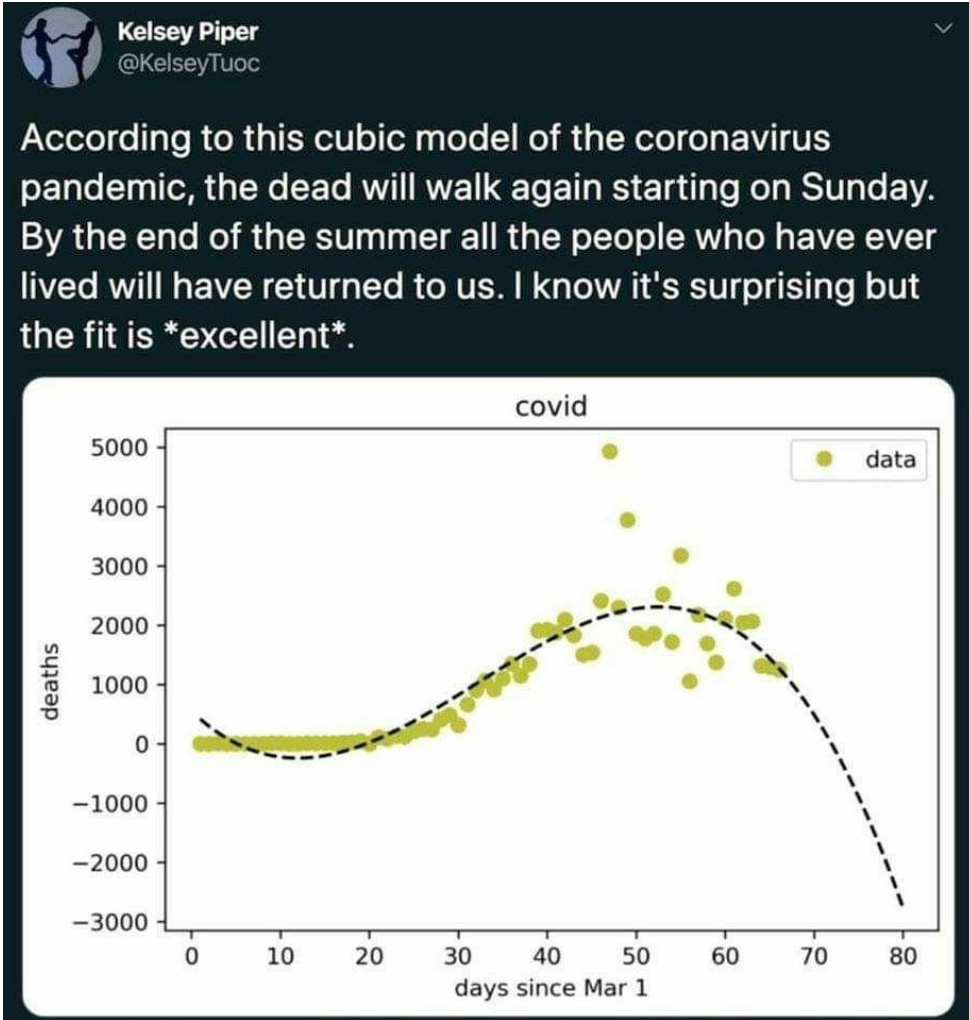
Как тестируют обучение с учителем?



# Overfit VS. Underfit



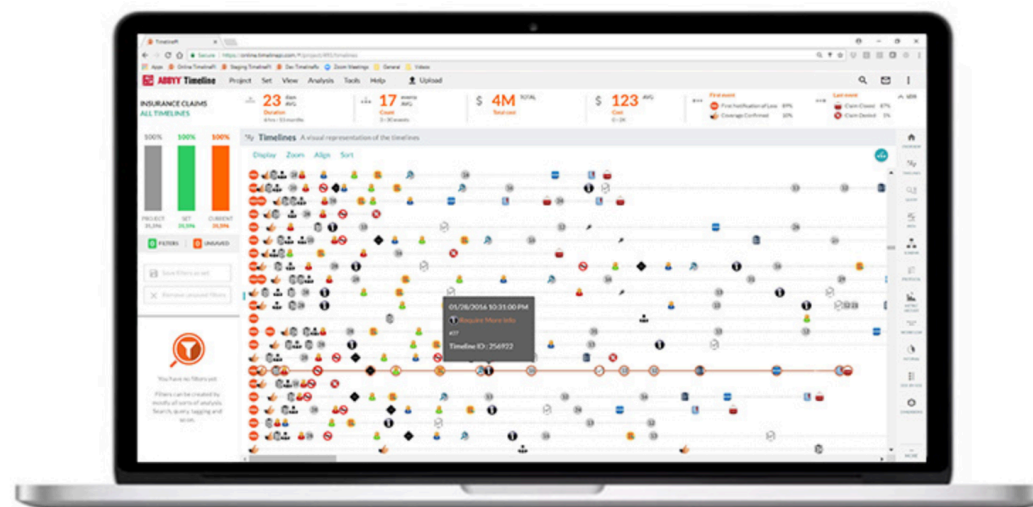
# Overfit VS. Underfit



## В сухом остатке

- AI — могила, ML — сила!
- Машинное обучение условно можно разделить на
  - обучение с подкреплением
  - обучение без учителя
  - обучение с учителем
- QA специалиста по данным — это не QA здорового человека.

# ABBYY Timeline



Зачем это клиенту?

**ABBYY®**





# Где получить входные данные?



Попросить DATA SET у разработчиков



Разметить самостоятельно?

# Прогнозирование событий

NR ALL TIMELINES

18 hrs - 20 days DURATION

6 - 16 events COUNT

6 - 14 events COUNT UNIQUE

4 hrs - 10 days TIME GAP

Invoice Received - 78% FIRST EVENT

Invoice LAST E...

Training started

## Forecast

Forecast



New

Open

Save

Save as

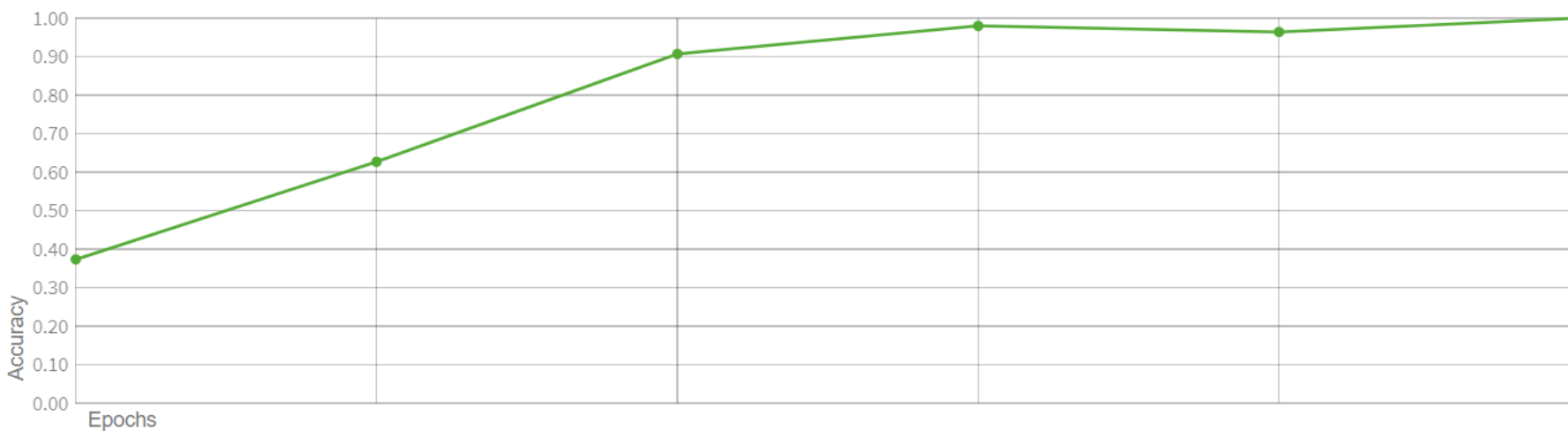
Configuration

Abort

Stop

Filter

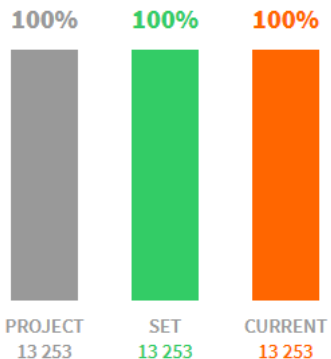
Status: Training round 6



Model Accuracy: %

Confusion Matrix:

|   | Actual |
|---|--------|
| 1 |        |



0 FILTERS | 0 UNSAVED

Save as set

Remove unsaved

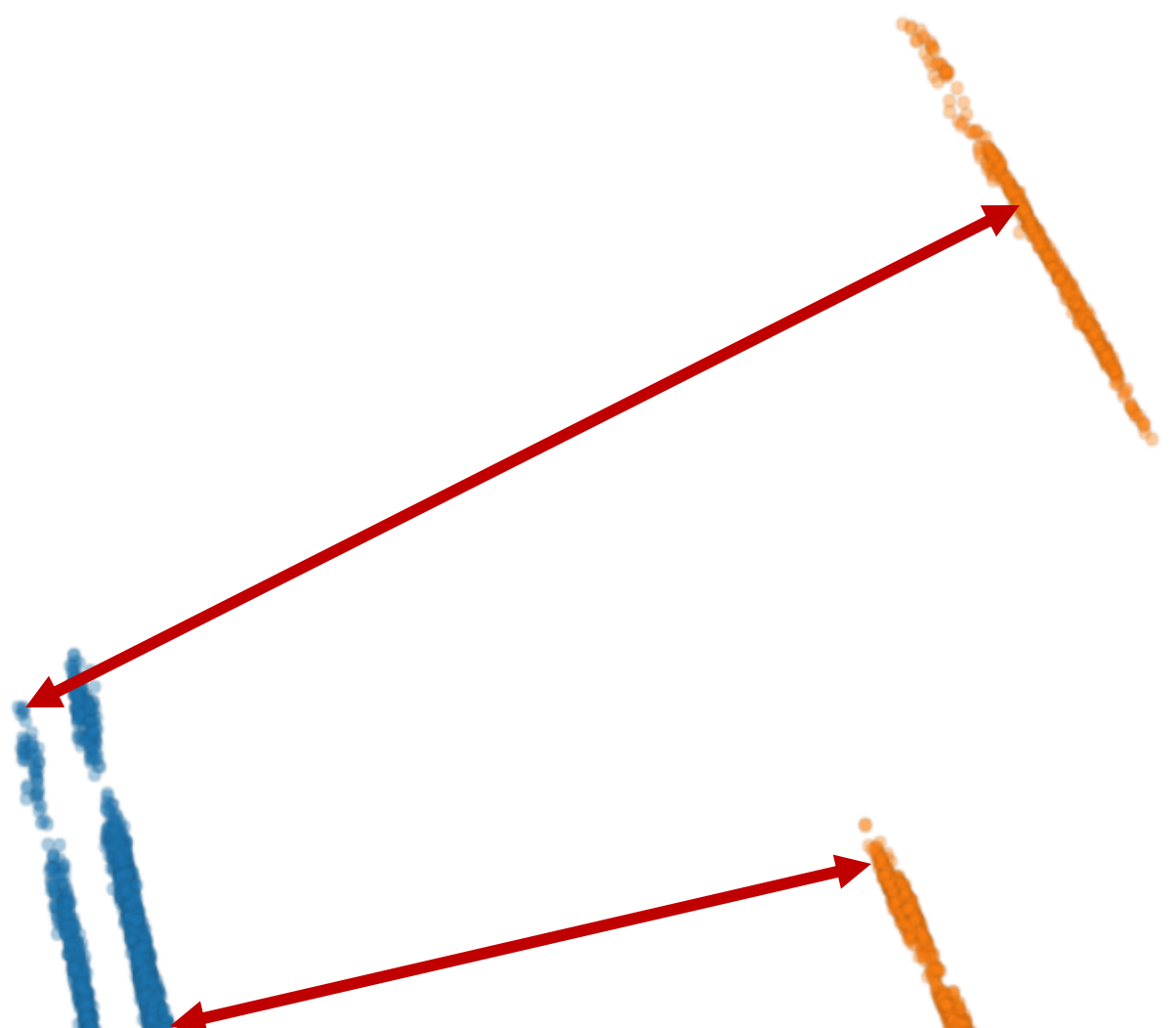


You have no filters yet

Filters can be created by mostly all sorts of analysis.

- OVERVIEW
- TIMELINES
- QUERY
- PATH
- SCHEMA
- PROTOCOL
- METRIC HISTORY
- WORKFLOW
- INTERVAL
- SIDE-BY-SIDE
- 18 MORE

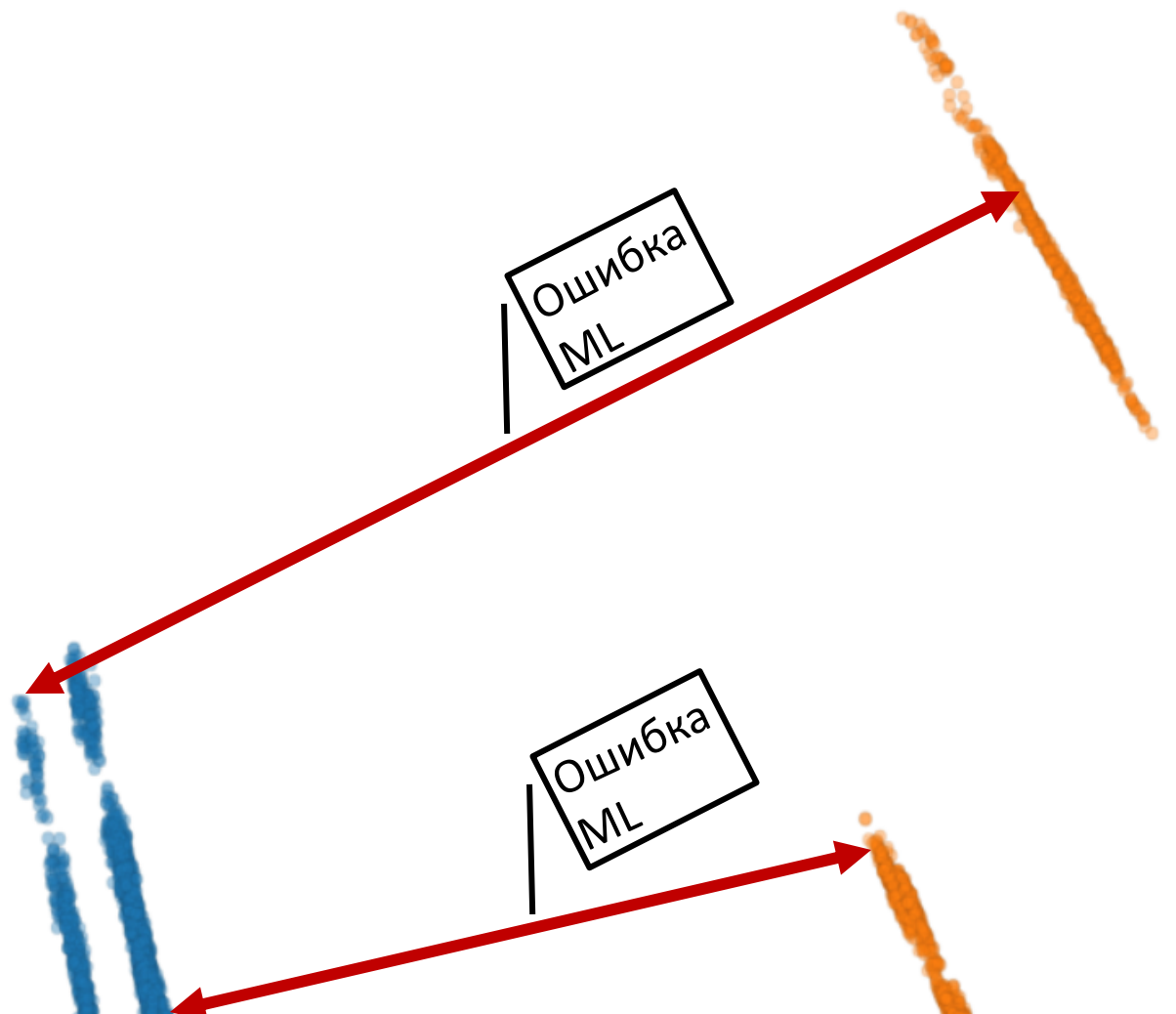
Timeline clusters:



 1



Timeline clusters:



Ошибка UI

1

# Как тестировать?



# А вы уверены, что результат обучения корректный?

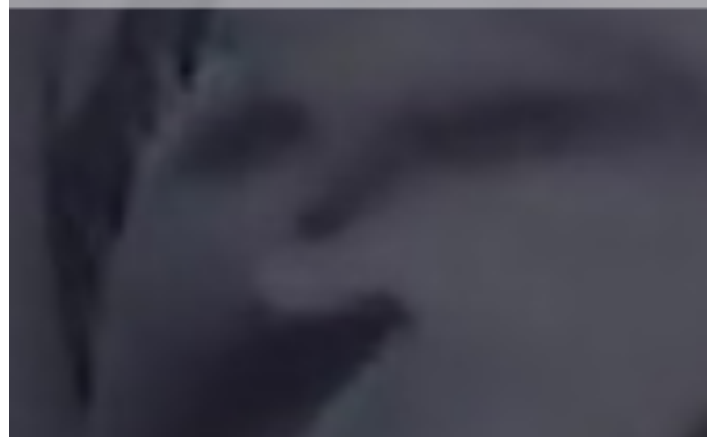


Данные переданные разработчиками  
подходили для тестирования?



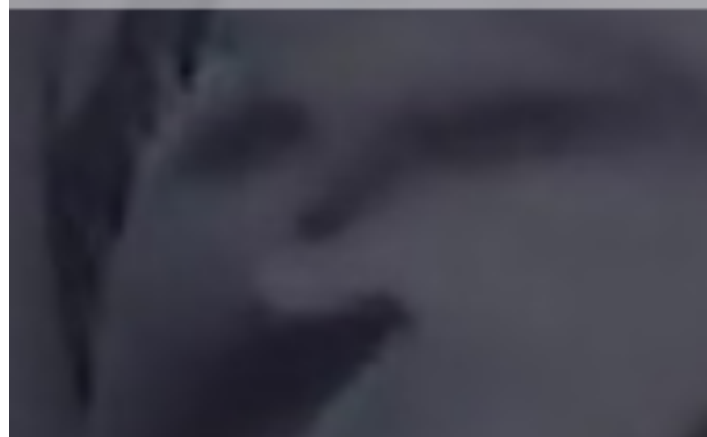
Вы точно понимаете как должен выглядеть  
корректный результат обучения?

Обнаруженные лица

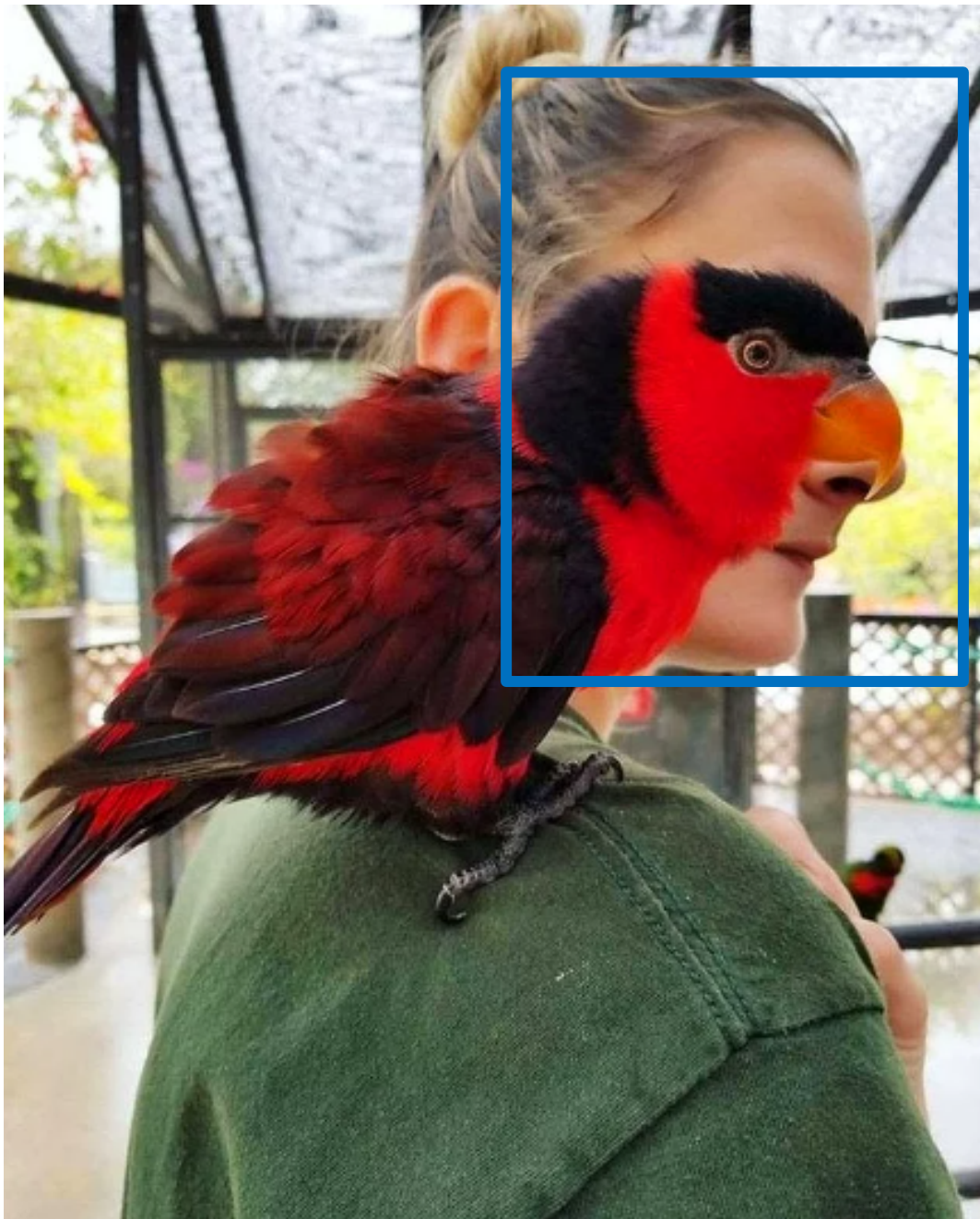


ТОЛЬКО ЧТО

Обнаруженные лица



ТОЛЬКО ЧТО



Обнаруженные лица

ТОЛЬКО ЧТО

Обнаруженные лица

ТОЛЬКО ЧТО

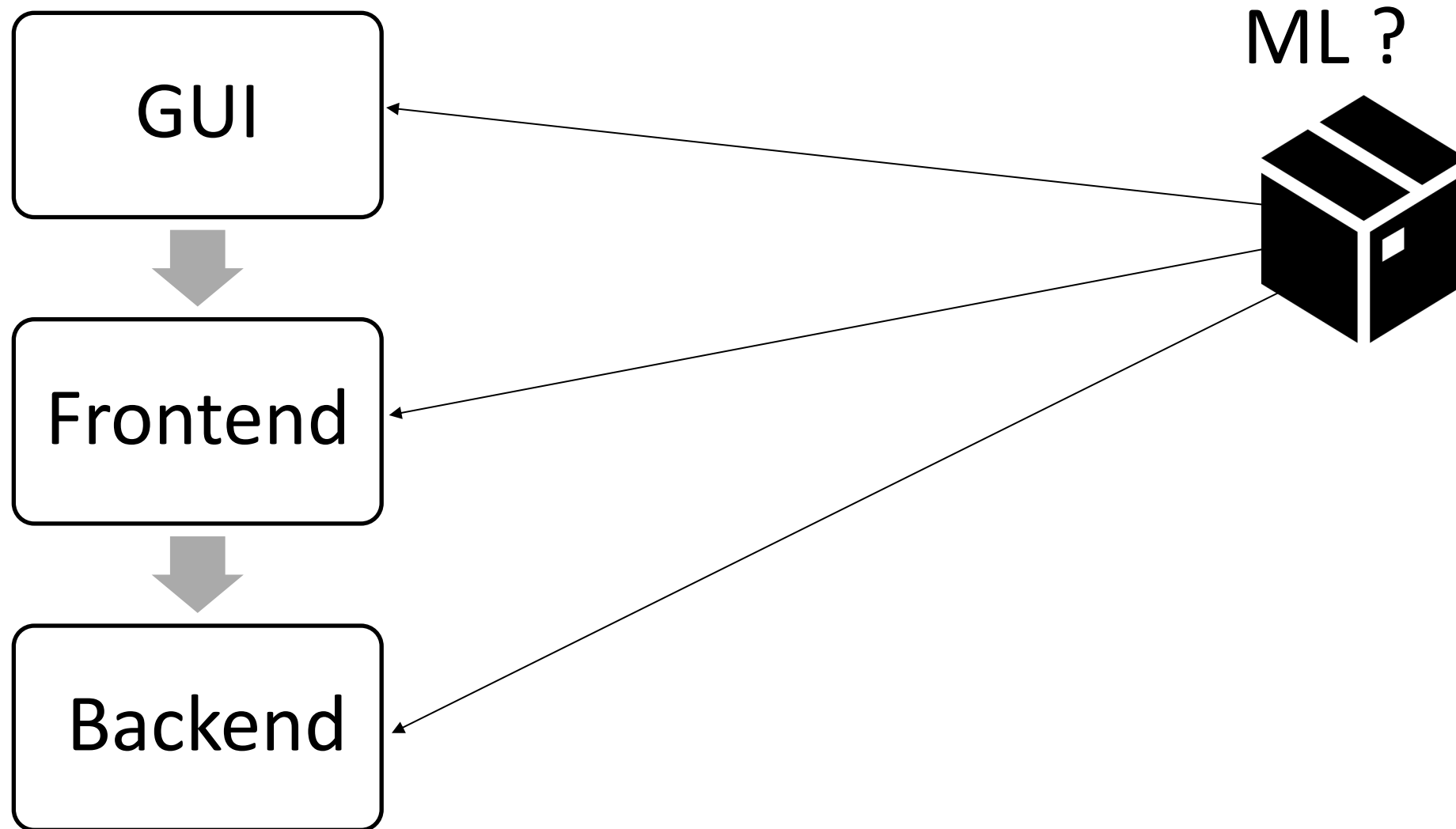


А вы уверены, что проблема в ML?

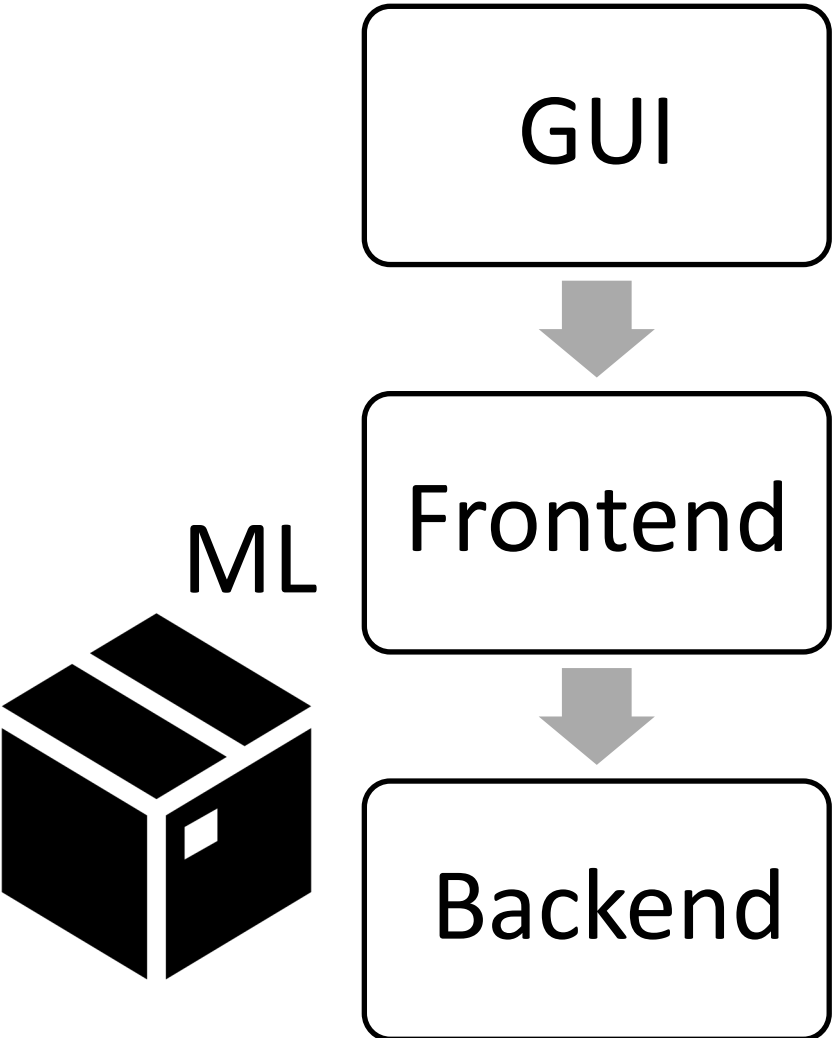
# Как упростить коммуникацию со специалистами по данным?

- Термины из мира ML
- Пройти обзорный курс по ML
- Какую задачу клиента ML решает в вашем продукте?
- Чем отличаются "реальные данные" от "данных в вакууме"
- А это вообще задача для ML?

# Когда тестировать ML?



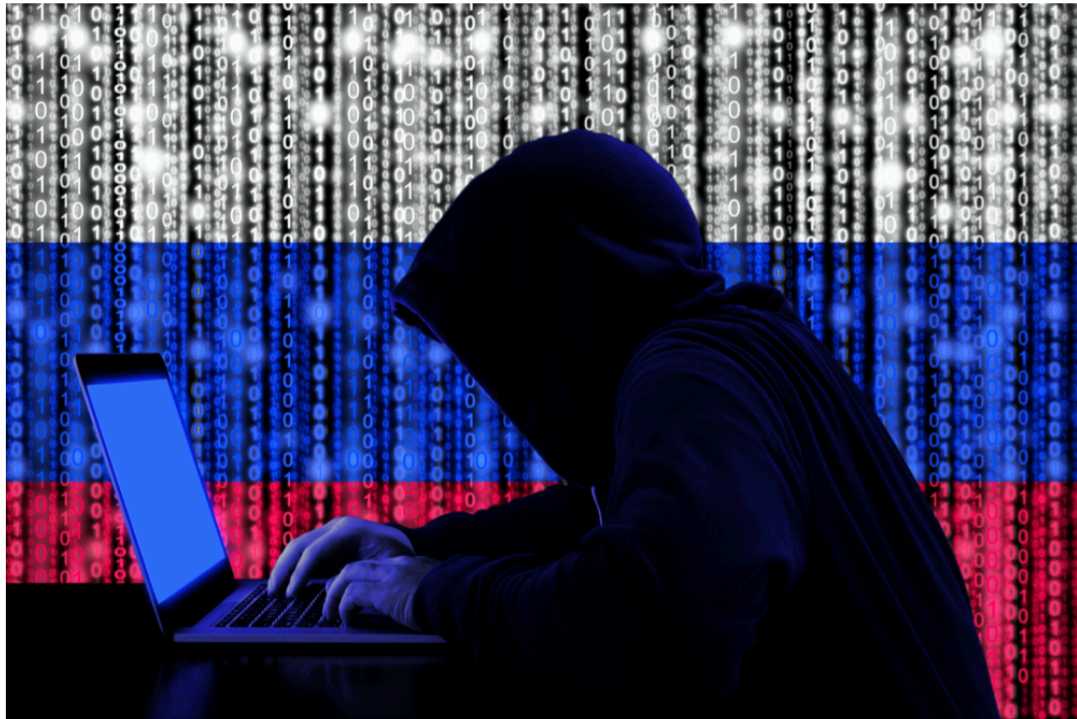
# Место ML в процессе тестирования



Чёрный ящик — это несколько "слоёв" сложности.



Международная команда — это непросто.



# Fail'ы разных уровней

1. UI
2. Термины из мира ML
3. Особенности работы AI
4. X-fail
5. 80 Level

Уровень 1:  
у меня нет знаний о ML



Forecast 

Nouveau

Ouvrir


Enregistrer

Enregistrer sous

Configuration

Commencer la formation

Réinitialiser les donn

 Filtrer



Forecast 

Nouveau    Ouvrir    Enregistrer    Enregistrer sous

 Filtrer

Configuration

Commencer la formation

Réinitialiser les donn

Ошибка UI

1.00  
0.90  
0.80  
0.70  
0.60  
0.50  
0.40  
0.30  
0.20  
0.10  
0.00

précision

époques

Уровень 2:  
знаю базовые термины ML

"SUBTITLE":

"Accuracy statistics of training by each epochs"

"SUBTITLE":

"Точность статистики обучения по каждой эре"

"SUBTITLE":

"Точность статистики обучения по каждой эре"

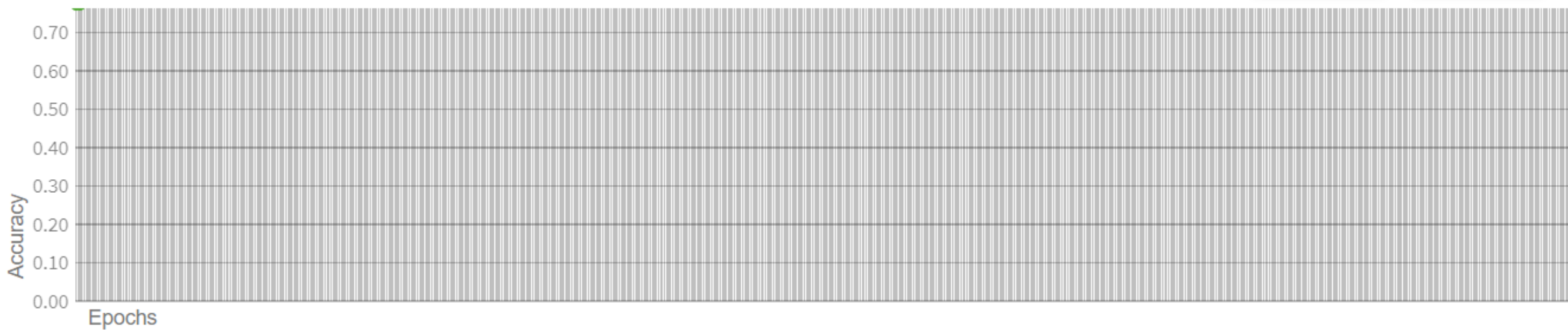
Некорректный  
термин

# Куда расти дальше?

- Как выбрать данные для проверки?
- Как проанализировать их не руками?

Уровень:  
знаю базовые принципы AI





Model Accuracy: 99.96%

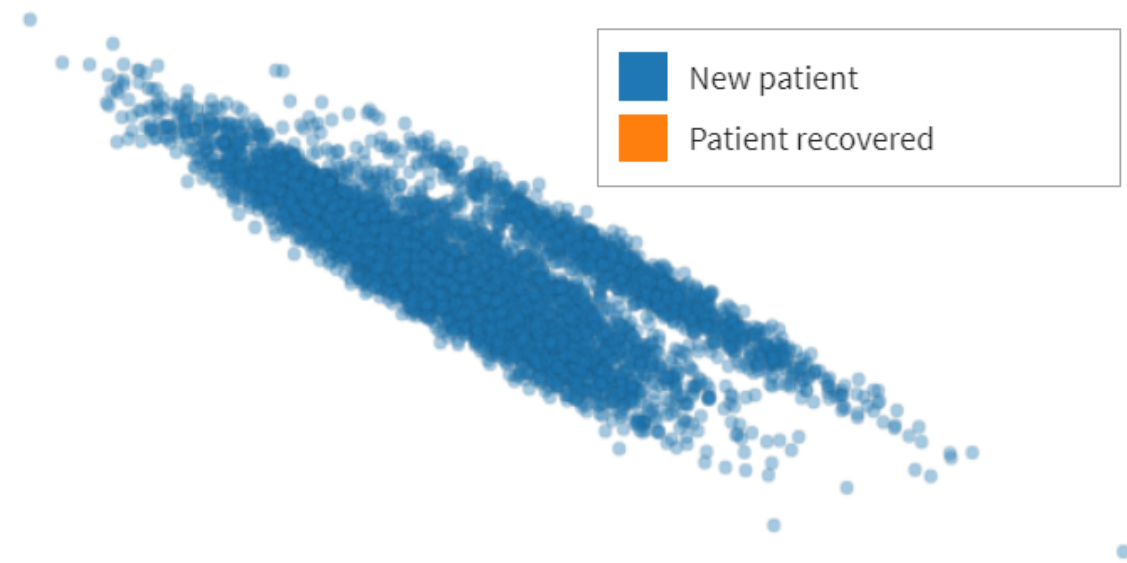
Confusion Matrix:

|           |                   | Actual      |                   |
|-----------|-------------------|-------------|-------------------|
|           |                   | New patient | Patient recovered |
| Predicted | New patient       | 22.892%     | 0.000%            |
|           | Patient recovered | 0.044%      | 77.064%           |

Event Importance

Event Importance

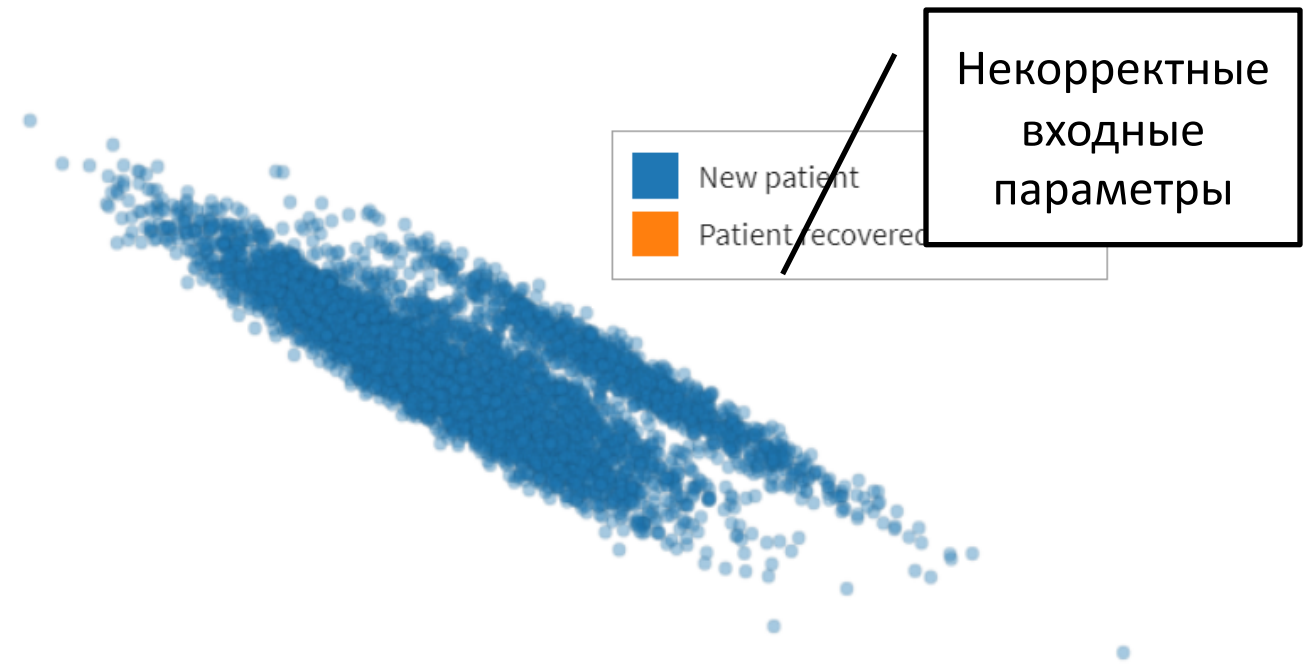
Timeline clusters:

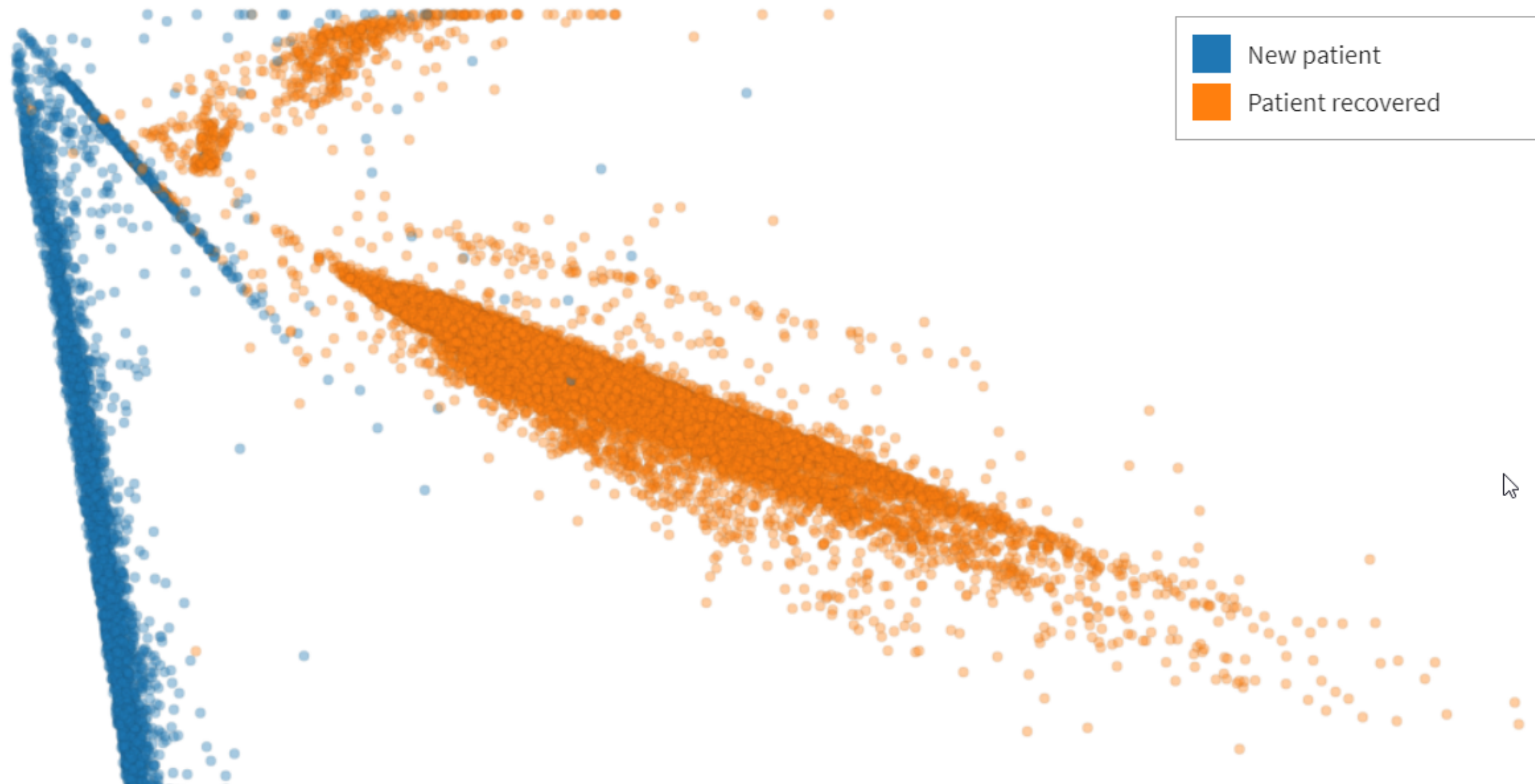


Event Importance

Event Importance

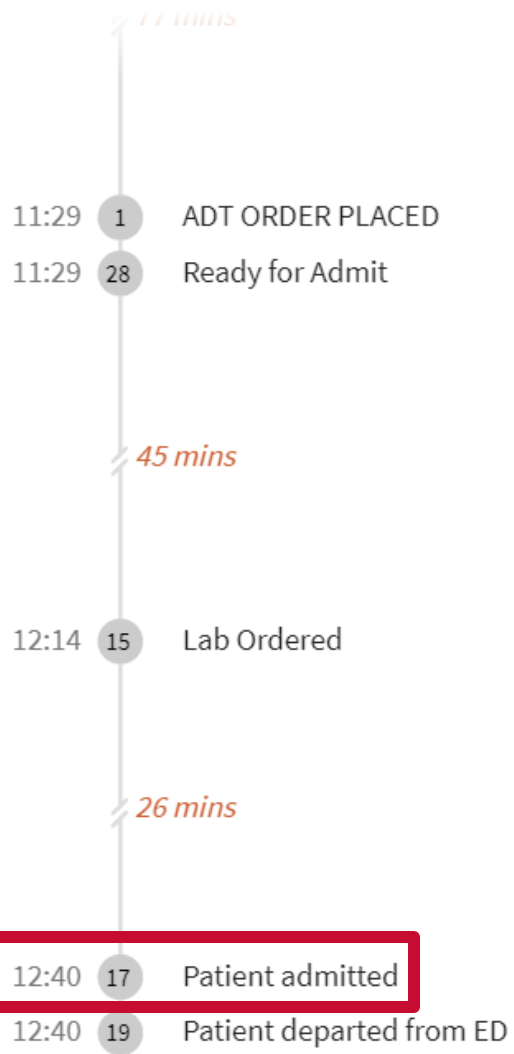
Timeline clusters:





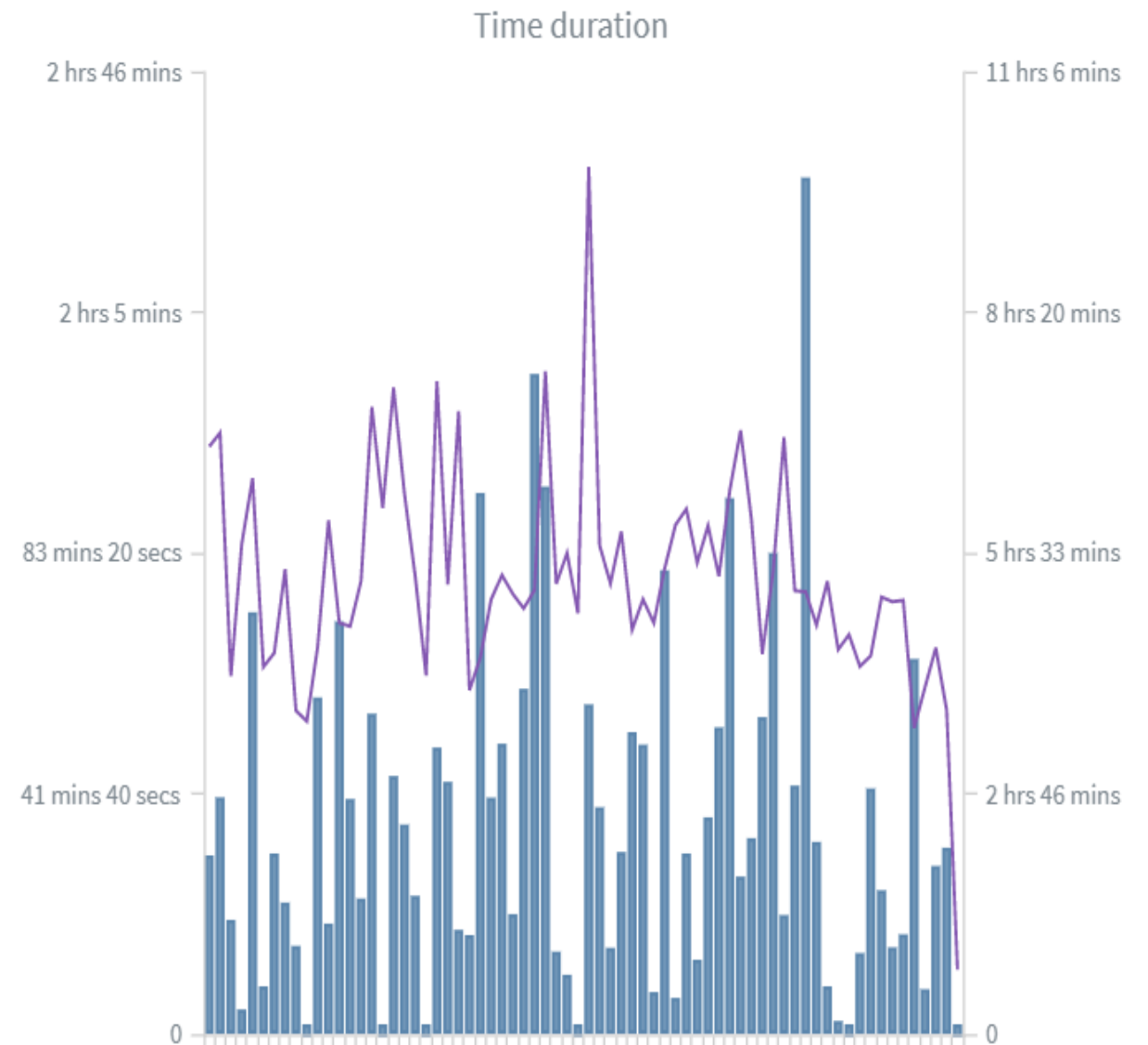
Legend:

- New patient
- Patient recovered



# Как должны выглядеть тестовые данные

- Посмотрите глазами на ваши данные
- Поймите что вы хотите получить
- График - ваш первый друг



# Уровень: Чего не делают DATA SCIENTIST'Ы

Training started LAST TRAINED

Forecast

f1 | 
 New | 
 Open | 
 Save | 
 Save as | 
 Configuration | 
 Start training | 
 Reset trained data | 
 Filter

Training of the model has failed. Please contact us.



Model Accuracy: 0.00%

Confusion Matrix:

|           |    | Actual |   |
|-----------|----|--------|---|
|           |    | q1     |   |
| Predicted | q1 |        | % |
|           |    |        |   |

- TIMELINES
- QUERY
- PATH
- SCHEMA
- PROTOCOL
- METRIC HISTORY
- WORKFLOW
- INTERVAL
- SIDE-BY-SIDE
- DIMENSIONS
- MORE



# Forecast

f1 New Open Save Save as Configuration Start training Reset trained data Filter

Training started Last event

Training of the model has failed. Please contact us.



Ожидаемое сообщение об ошибке

Model Accuracy: 0.00%

Confusion Matrix:

|           |    | Actual |  |
|-----------|----|--------|--|
|           |    | q1     |  |
| Predicted | q1 | %      |  |
|           |    |        |  |

QUERY

RTH

SCHEMA

PROTOCOL

METRIC HISTORY

WORKFLOW

INTERVAL

SIDE-BY-SIDE

DIMENSIONS

MORE

Уровень:  
80 Level

# Куда расти дальше?

- Как просить помощи у коллег и когда это критически важно?
- Метрики

0 - 35 days DURATION

1 - 102 events COUNT

1 - 30 events COUNT UNIQUE

0 - 35 days TIME GAP

Patient arrival FIRST EVENT

Loaded your latest viewed configuration for: "Forecast".

### Forecast

Need room



New

Open

Save

Save as

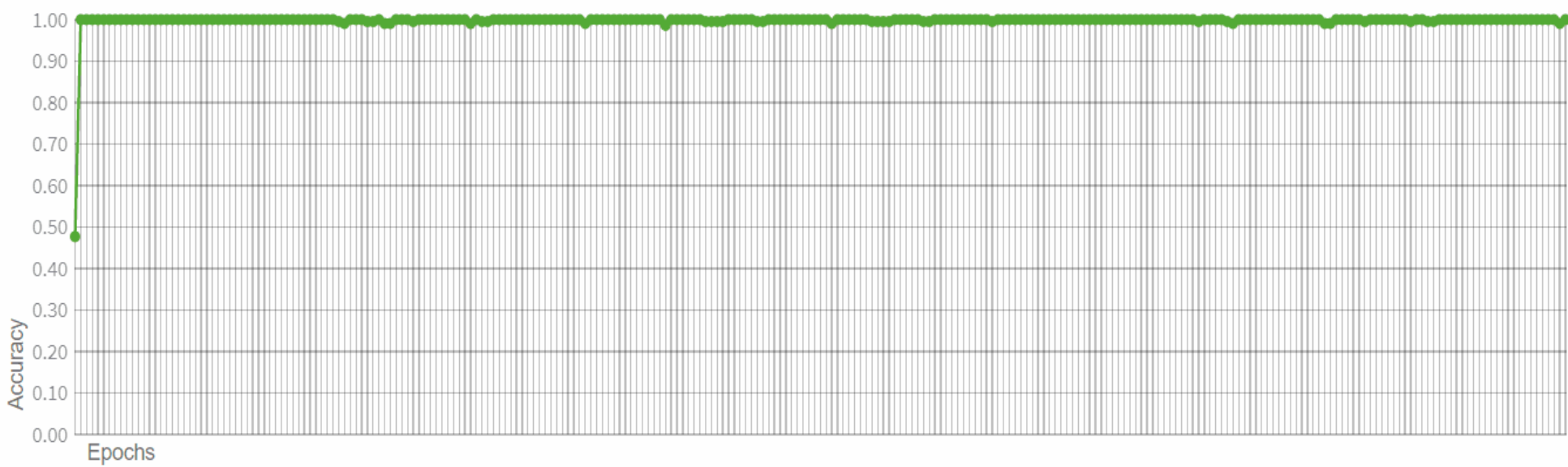
Configuration

Abort

Stop

Filter

Status: Training round 261



Model Accuracy: 100.00%

Confusion Matrix:

|           | Actual          |           |
|-----------|-----------------|-----------|
|           | Patient arrived | Need room |
| Predicted |                 |           |

100%  
100%

0% MET 472  
CURRENT 23 472

0 UNSAVED

et

unsaved



no filters yet

created by  
ts of analysis.

- OVERVIEW
- TIMELINES
- QUERY
- PATH
- SCHEMA
- PROTOCOL
- METRIC HISTORY
- WORKFLOW
- INTERVAL
- SIDE-BY-SIDE
- MORE

## Вместо выводов

- Спрос на тестирование ML будет только расти
- QA в ML требует специальных знаний
- Вам нужна прямая коммуникация со специалистами по анализу данных
- Вам нужны специальные знания, и мало их не бывает
- Научитесь раскладывать ошибки по полочкам

**ABBYY®**

Спасибо!

---

