

Дизайн систем машинного обучения

6. Оценка качества модели

План курса

- 1) Практическое применение машинного обучения
- 2) Основы проектирования ML-систем
- 3) Обучающие данные
- 4) Подготовка и отбор признаков
- 5) Выбор модели, разработка и обучение модели
- 6) Оценка качества модели — Вы находитесь здесь**
- 7) Развертывание систем
- 8) Диагностика ошибок и отказов ML-систем
- 9) Мониторинг и обучение на потоковых данных
- 10) Жизненный цикл модели
- 11) Отслеживание экспериментов и версионирование моделей
- 12) Сложные модели: временные ряды, модели над графами
- 13) Непредвзятость, безопасность, управление моделями
- 14) ML инфраструктура и платформы
- 15) Интеграция ML-систем в бизнес-процессы

Система в целом

- Интерфейсы (API, GUI etc.)
- Данные
- ML-модели
- Инфраструктура // обсудим на других лекциях
- Оборудование // обсудим на других лекциях

Качество: Интерфейсы

- API
 - Документация: OpenAPI, самодокументируемые API → →
 - Надежность: логгирование и мониторинг
 - Доступность: потребители смогут подключиться
 - Внутренние сети, старые протоколы, другие операционные системы, медленная сеть
- GUI
 - Удобство в использовании: usability →
 - Надежность
 - Доступность : потребители смогут подключиться
 - Мобильный клиент, старые браузеры, медленная сеть

Пример: FastAPI docs

- <https://data.nav.no/digdir-api/openapi.json>
- <https://fastapi.tiangolo.com/>
- <https://swagger.io/specification/>

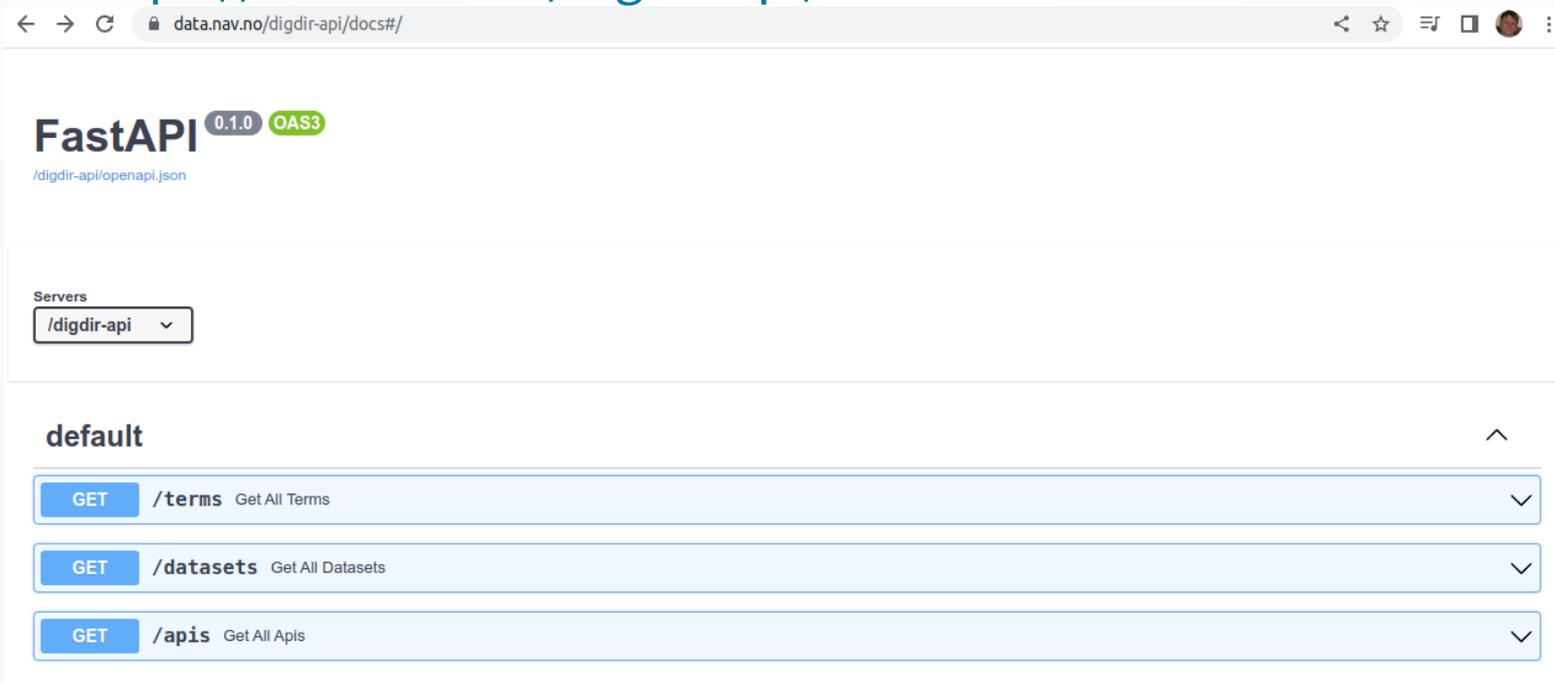


A screenshot of a web browser displaying an OpenAPI JSON specification. The address bar shows the URL `data.nav.no/digdir-api/openapi.json`. The page content is a JSON object representing the API specification, including information about the API version, servers, and various endpoints like `/terms`, `/datasets`, and `/apis`.

```
{
  "openapi": "3.0.2",
  "info": {
    "title": "FastAPI",
    "version": "0.1.0"
  },
  "servers": [
    {
      "url": "/digdir-api"
    }
  ],
  "paths": {
    "/terms": {
      "get": {
        "summary": "Get All Terms",
        "operationId": "get_all_terms_terms_get",
        "responses": {
          "200": {
            "description": "Successful Response",
            "content": {
              "text/turtle": {
                "schema": {
                  "type": "string"
                }
              }
            }
          }
        }
      }
    },
    "/datasets": {
      "get": {
        "summary": "Get All Datasets",
        "operationId": "get_all_datasets_datasets_get",
        "responses": {
          "200": {
            "description": "Successful Response",
            "content": {
              "text/turtle": {
                "schema": {
                  "type": "string"
                }
              }
            }
          }
        }
      }
    },
    "/apis": {
      "get": {
        "summary": "Get All Apis",
        "operationId": "get_all_apis_apis_get",
        "responses": {
          "200": {
            "description": "Successful Response",
            "content": {
              "text/turtle": {
                "schema": {
                  "type": "string"
                }
              }
            }
          }
        }
      }
    }
  }
}
```

Пример: FastAPI docs

- <https://data.nav.no/digdir-api/docs>



The screenshot shows a web browser displaying the FastAPI documentation for the digdir-api. The page title is "FastAPI 0.1.0 OAS3" with a link to "/digdir-api/openapi.json". Below the title, there is a "Servers" section with a dropdown menu showing "/digdir-api". The main content area is titled "default" and lists three endpoints:

- GET** `/terms` Get All Terms
- GET** `/datasets` Get All Datasets
- GET** `/apis` Get All Apis

Пример: FastAPI docs

The screenshot shows the FastAPI documentation interface for the `GET /terms` endpoint. The page is titled "Parameters" and "Responses". The "Parameters" section indicates "No parameters". The "Responses" section shows a table with one entry: a 200 status code for a "Successful Response" with "No links". Below the table, there is a "Media type" dropdown menu set to "text/turtle" and a note "Controls Accept header.". Underneath, there are links for "Example Value" and "Schema". The "Example Value" section shows a dark box containing the text "string".

GET /terms Get All Terms

Parameters Try it out

No parameters

Responses

Code	Description	Links
200	Successful Response	No links

Media type

text/turtle

Controls Accept header.

Example Value | Schema

```
string
```

Качество: Данные

- <https://greatexpectations.io/>



Expectations

Expectations are assertions for data. They are the workhorse abstraction in Great Expectations, covering all kinds of common data issues.

Expectations are declarative, flexible and extensible. They provide a rich vocabulary for data quality.

[Check out the Expectation Gallery](#)

- `expect_column_values_to_not_be_null`
- `expect_column_values_to_match_regex`
- `expect_column_values_to_be_unique`
- `expect_column_values_to_match_strftime_format`
- `expect_table_row_count_to_be_between`
- `expect_column_median_to_be_between`

Пример: Great Expectations Validation

great_expectations Home / Validations / raw_residential_data.warning / raw_residential_data_failures / 2020-11-24T02:24:42.888548+00:00

Actions

Validation Filter:

Show All Failed Only

How to Edit This Suite

Show Walkthrough

Table of Contents

- Overview
- Table-Level Expectations
- BATHRM
- OBJECTID
- SSL

Status	Expectation	Observed Value																																										
✓	values must never be null.	100% not null																																										
	values must always be greater than or equal to 1 and less than or equal to 11 characters long.	renderer_type="renderer.prescriptive"																																										
	227 unexpected values found. ≈0.2118% of 107154 total rows.	renderer_type="renderer.diagnostic.unexpected_statement"																																										
	<table border="1"><thead><tr><th>Unexpected Value</th><th>Count</th></tr></thead><tbody><tr><td>BBBBBBBBBBBBBBBB</td><td>69</td></tr><tr><td>ABABDEDEGEGE</td><td>58</td></tr><tr><td>CCCCBBBEEDEE</td><td>32</td></tr><tr><td>AAAABBBBCCCC</td><td>21</td></tr><tr><td>5881S 0004</td><td>5</td></tr><tr><td>5883 0062</td><td>5</td></tr><tr><td>5890 0123</td><td>4</td></tr><tr><td>5890 0134</td><td>4</td></tr><tr><td>0786 0820</td><td>3</td></tr><tr><td>1253 0139</td><td>3</td></tr><tr><td>2125 0024</td><td>3</td></tr><tr><td>2231 0007</td><td>3</td></tr><tr><td>5890 0122</td><td>3</td></tr><tr><td>0237 0822</td><td>2</td></tr><tr><td>0238 0800</td><td>2</td></tr><tr><td>0306 0020</td><td>2</td></tr><tr><td>0366 0022</td><td>2</td></tr><tr><td>0694 0036</td><td>2</td></tr><tr><td>0870 0122</td><td>2</td></tr><tr><td>0989 0071</td><td>2</td></tr></tbody></table>	Unexpected Value	Count	BBBBBBBBBBBBBBBB	69	ABABDEDEGEGE	58	CCCCBBBEEDEE	32	AAAABBBBCCCC	21	5881S 0004	5	5883 0062	5	5890 0123	4	5890 0134	4	0786 0820	3	1253 0139	3	2125 0024	3	2231 0007	3	5890 0122	3	0237 0822	2	0238 0800	2	0306 0020	2	0366 0022	2	0694 0036	2	0870 0122	2	0989 0071	2	renderer_type="renderer.diagnostic.unexpected_table"
Unexpected Value	Count																																											
BBBBBBBBBBBBBBBB	69																																											
ABABDEDEGEGE	58																																											
CCCCBBBEEDEE	32																																											
AAAABBBBCCCC	21																																											
5881S 0004	5																																											
5883 0062	5																																											
5890 0123	4																																											
5890 0134	4																																											
0786 0820	3																																											
1253 0139	3																																											
2125 0024	3																																											
2231 0007	3																																											
5890 0122	3																																											
0237 0822	2																																											
0238 0800	2																																											
0306 0020	2																																											
0366 0022	2																																											
0694 0036	2																																											
0870 0122	2																																											
0989 0071	2																																											
✗		renderer_type="renderer.diagnostic.observed_value" ≈0.21184% unexpected																																										

Пример: Great Expectations DataDoc

ratings

Overview

id

user_id

movie_id

rating

rated_at

movie_id (type: int)

Properties

Distinct (n)	1682
Distinct (%)	1.7%
Missing (n)	0
Missing (%)	0.0%

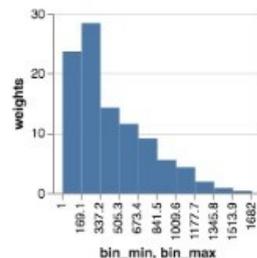
Quantiles

0.05	30
Q1	175
Median	322
Q3	631
0.95	1074

Statistics

Mean	425.53
Minimum	1.00
Maximum	1682.00

Histogram



Example values

29 40 51 86 222 242 257
265 274 302 346 377 387
451 465 474 785 1014 1042
1184

Пример: Паттерн «Предохранитель»

Data Source Issues

- Table inconsistencies
 - a. Illegitimate values
 - b. Missing values
 - c. Duplicate Primary keys
- Hard deletes
- Bulk inserts
- Missing updates to CDC column

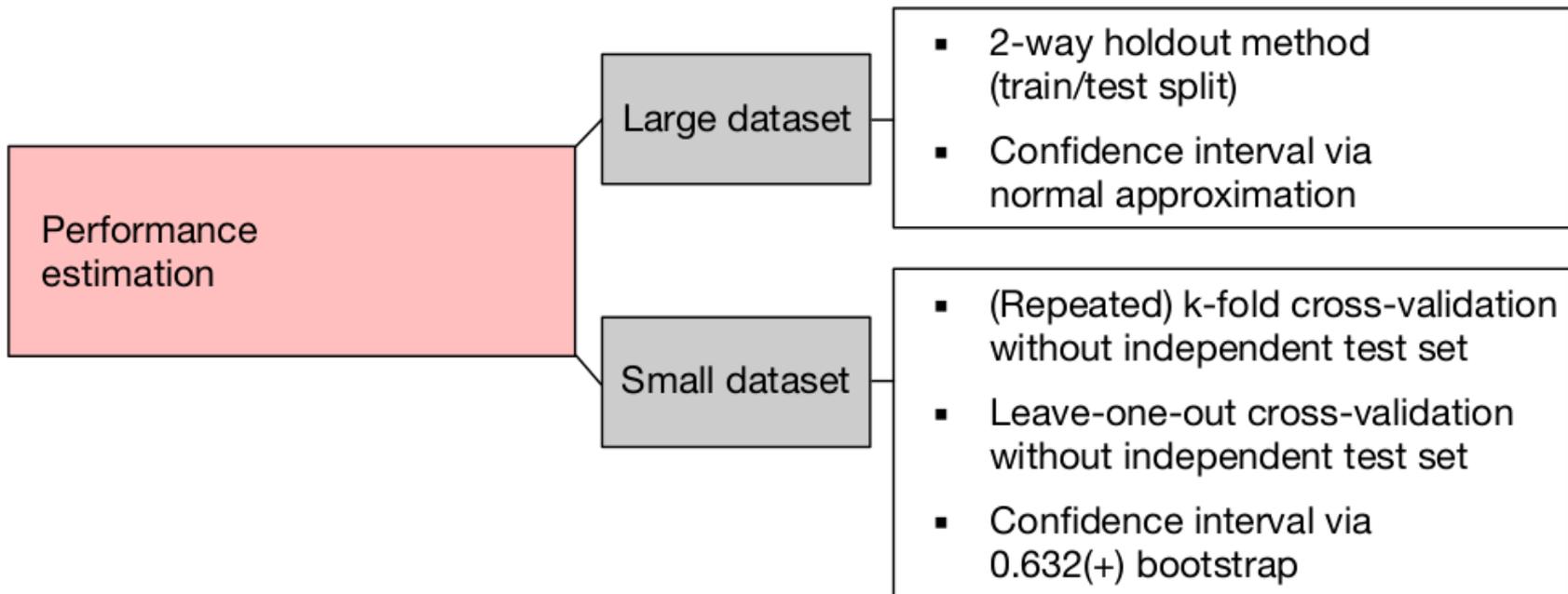
Data Ingestion Issues

- Uncoordinated upstream changes
 - a. Volume of data
 - b. Change in schema
 - c. Change in meaning of data
 - d. Upgrade of platform
- No CDC for large tables leading to delayed availability
- Errors in ETL logic
- Timezone inconsistencies
- Duplicate or null records due to ingestion errors

Referential Integrity Issues

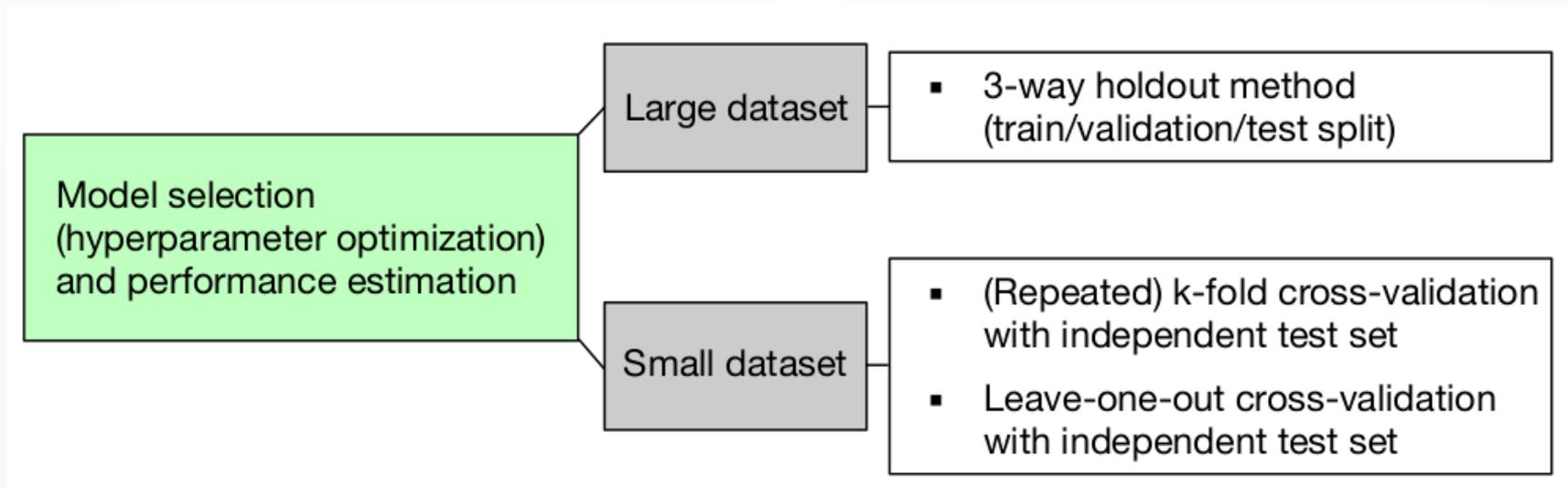
- Data elements have different data types and/or meaning in different sources
- Inconsistent data element enums
- Heuristic ID Correlation
- Uncoordinated schema changes
- Dropped updates across data sources

Оценка качества ML-модели



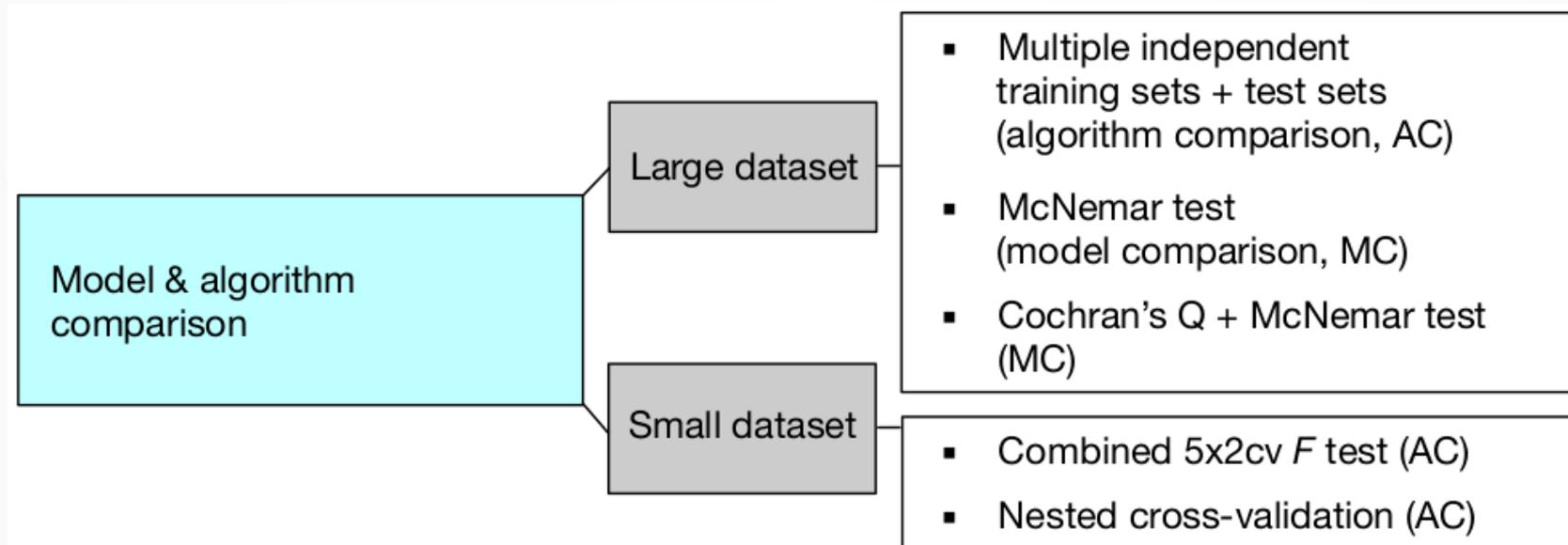
<https://arxiv.org/abs/1811.12808> must read

Оценка качества ML-модели



<https://arxiv.org/abs/1811.12808> must read

Оценка качества ML-модели



<https://arxiv.org/abs/1811.12808> must read

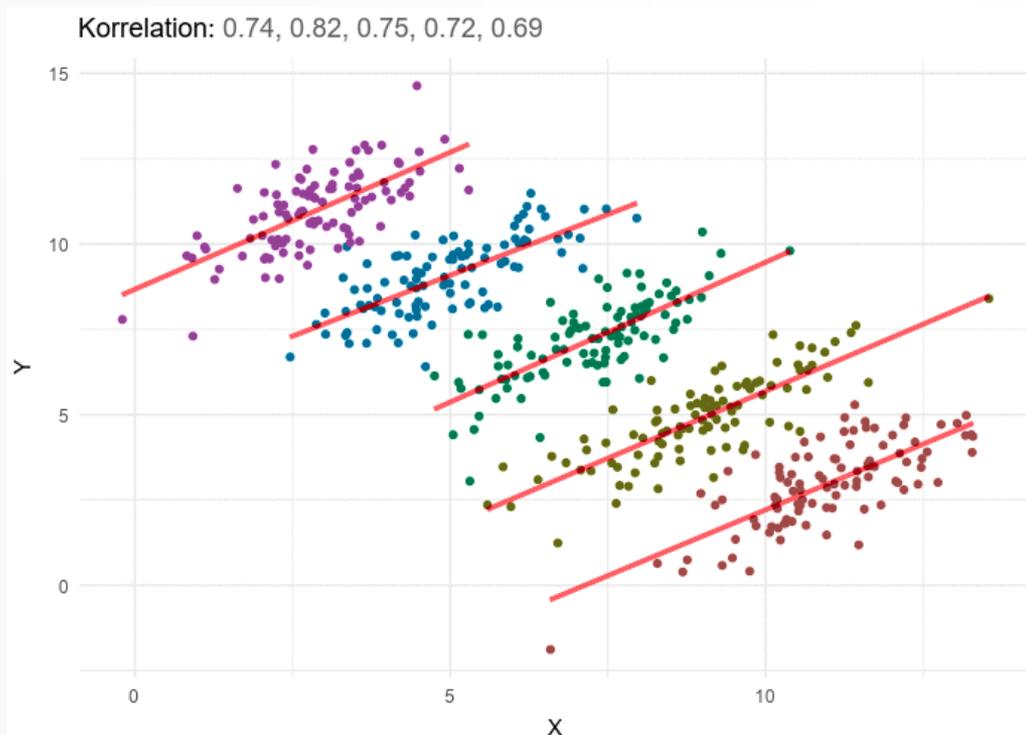
С чем сравнивать

- Случайный бейзлайн
- В соответствии с распределением меток
- Самый частый класс
- Простые эвристики
- Качество оценки человеком
- Существующие решения
- См [sample_quality.pdf](#)

Тесты устойчивости

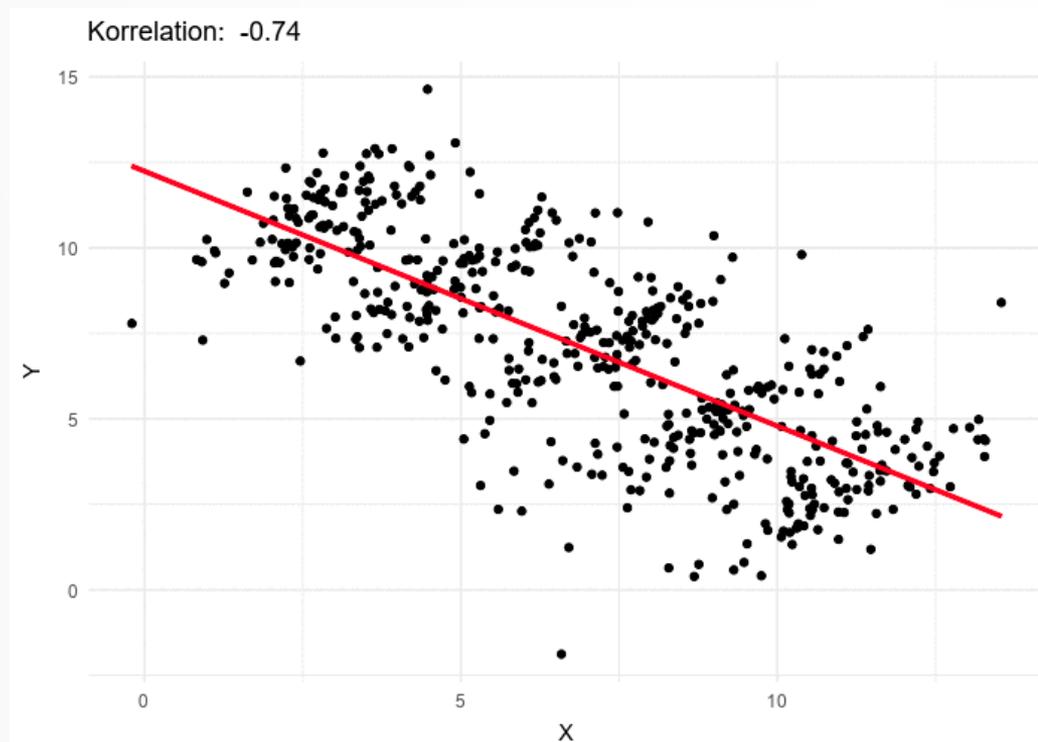
- Perturbation tests
 - Насколько чувствительна модель к шуму?
- Invariant test
 - Что не должно поменять предсказание модели?
- Directional Expectations test
 - Что должно менять предсказание модели?
 - Например
 - в кредитном скоринге: меньше доход → меньше скор
 - Медосмотр: чем старше, тем хуже здоровье

Парадокс Симпсона



https://en.wikipedia.org/wiki/Simpson%27s_paradox

Парадокс Симпсона



https://en.wikipedia.org/wiki/Simpson%27s_paradox

Оценка на подвыборках

- Можно найти кластеры и оценить качество работы на них
- Subgroup Discovery Algorithms: A Survey and Empirical Evaluation
- Automated Data Slicing for Model Validation

см [sample_quality.pdf](#)

Дополнительные материалы

- [Reducing Pipeline Debt With Great Expectations](#) (blog)
- [Effective testing for machine learning systems](#) (blog)
- [Microsoft Error Analysis Toolkit](#) (tool, tutorial)
- [RAI Toolbox](#) (video)