Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Data Management for Digital Health
Revision of Exercise 1

Borchert, Dr. Schapranow

Data Management for Digital Health
Winter 2023

Exercise I
Topics

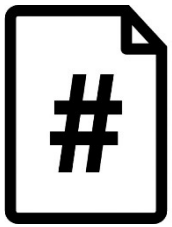
- Actors in Healthcare
- Iterative Machine Learning Design Process
- Categories of Data
- Health Data Exchange

Exercise I
Key Stats

Hasso Plattner Institut

25 Questions 50 Points	27 Students 27 Passed	Average score 44.3 / 88.6%	Average time 79.3 min

<< 3h



Q3: Categories of Data

Please select all appropriate statements for health systems as discussed in class:

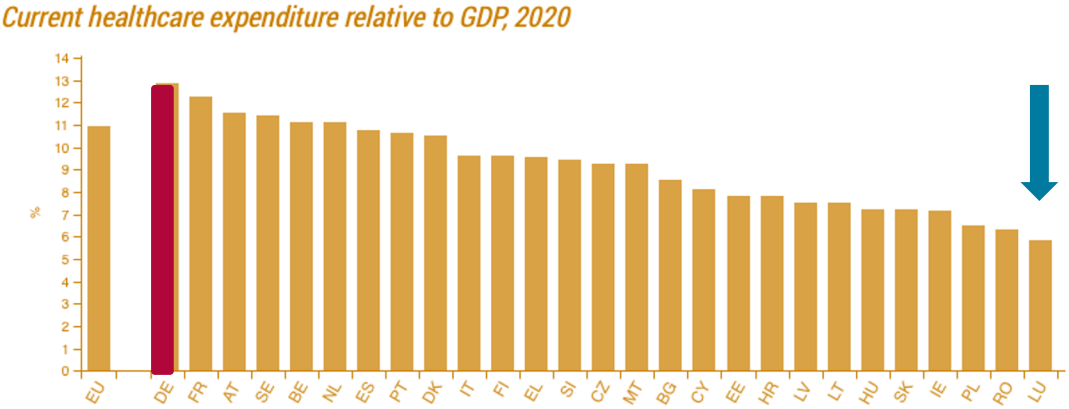
- ✓ There are different types of health insurances for citizens available, e.g., statutory and private health insurances.
- ✗ The costs per capita and year are a strong indicator for the quality of health care services.
- ✗ The government covers all costs for healthcare services of citizens.
- ✗ Health insurances define regulations and legal frameworks for national healthcare strategies.

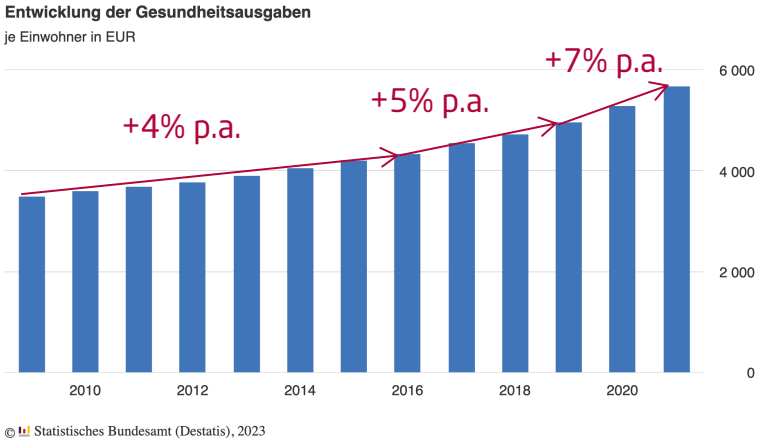
Frequently missed

Frequent incorrect answer

Facts You Should Know
Costs for Healthcare

Hasso Plattner Institut





- Germany 5,699 EUR per capita resp. 13.2 % of GDP in 2021

Introduction to Digital Health

Data Management for Digital Health, Winter 2023

OECD Health Statistics 2023, Eurostat Database, WHO, Global Health Expenditure Database, Destatis
https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare_expenditure_statistics

Federal Ministry of Health (BMG)

- Development of laws, legal frameworks, and regulations
- Supervision of:
  - Statutory health insurances,
  - Nursing care insurance,
  - Health protection,
  - Health professions, and
  - Pharmaceuticals & medical devices.

https://www.bundesgesundheitsministerium.de/ministerium/aufgaben-und-organisation/aufgaben.html

Ressortforschung des Bundesministeriums für Gesundheit

Das Bundesministerium für Gesundheit (BMG) bereitet regelmäßig weitreichende politische Entscheidungen vor, die einen direkten Einfluss auf das alltägliche Leben der Bürgerinnen und Bürger haben.

▶ weiterlesen

Q4: Please select the correct statements as discussed in class.

Hasso Plattner Institut

- Information is received data.
- Data are information symbolized as material-energetic pattern
- Knowledge is stored experience.
- Information can never be forgotten once is was recorded.

Frequently missed

Frequent incorrect answer

Terms You Should Know

- **Data** := Information symbolized as material-energetic pattern, a.k.a. message
- **Information** := Something knowable, i.e. something we could know
  - **Perceiving** := Perceived information from someone else
  - **Interpretation** := Received information
  - **Processing** := Derived information
  - **Forgottenness** := Lost information
- **Knowledge** := Stored experience

Wendt, Siegfried. Nichtphysikalische Grundlagen der Informationstechnik: Interpretierte Formalismen. Springer-Verlag, 1991.

Q8: Let us consider the data management process for digital health data as introduced in class. Which of the following process steps are part of it?

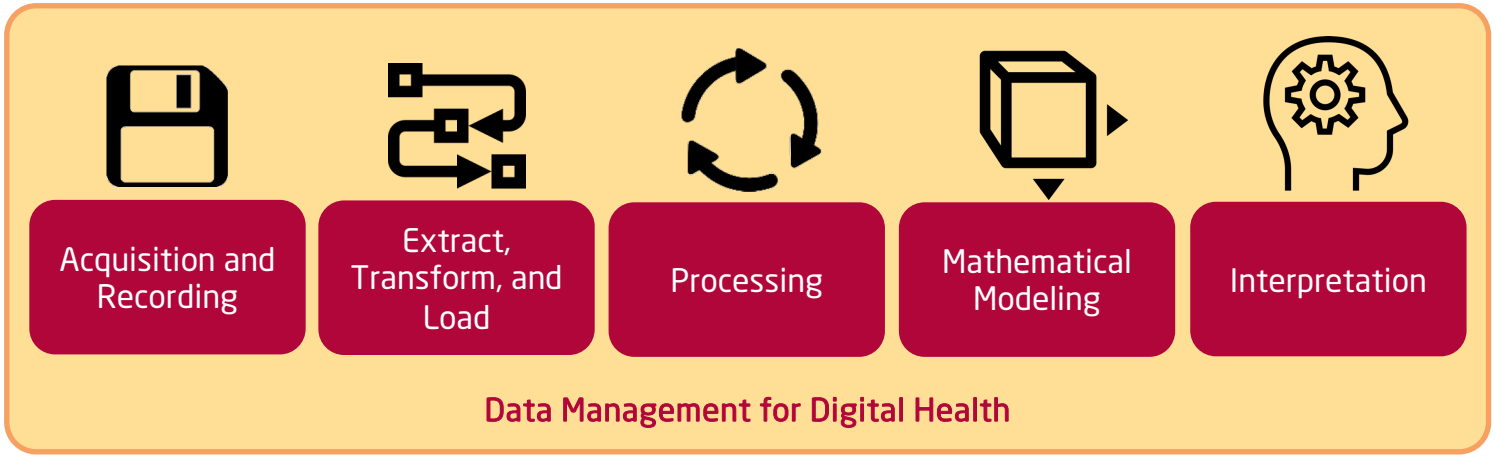
Hasso Plattner Institut

- ✓ The so-called ETL process, i.e., extract, transform, and load of data.
- ✓ Interpretation of analysis results by a subject-matter expert, e.g., medical professionals, to assess medical meaningfulness of the analysis results.
- ✗ Identification of the potential end user within the corresponding hospital department.
- ✗ Selection of an adequate clinical ontology to acquire data.

Frequently missed

Frequent incorrect answer

Data Management
Process Perspective



Q11: Consider an algorithm that, given a medical image, predicts a segmentation mask highlighting a tumor region. This is an example for:

- ✗ Regression
- ✓ Structured Prediction
- ? Clustering
- ✗ Reinforcement Learning

Frequently missed

Frequent incorrect answer

Most common problem settings in Machine Learning

Hasso Plattner Institut

Supervised Learning (Labels available for training)

Classification

Categorical output

e.g. x ∈ Fruits, y ∈ {"apple", "orange"}

f()= "apple"

f()= "orange"

Regression

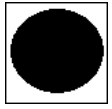
Continuous output

e.g.: x ∈ Fruits, y ∈ $\mathbb{R}_+$ ≙ t until ripe)

f () = 12 days

Structured Prediction

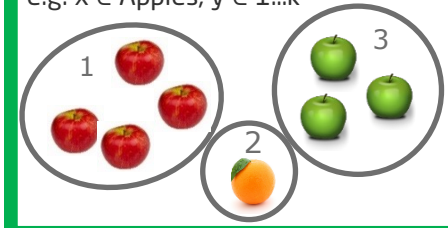
e.g. x ∈ $\mathbb{R}^{w \times h \times d}$, y ∈ $\mathbb{R}^{w \times h}$ ≙ pixels

f() = 

Unsupervised Learning (No labels during training)

Clustering

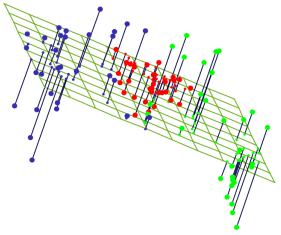
e.g. x ∈ Apples, y ∈ 1...k



Dimensionality reduction

x ∈ $\mathbb{R}^d$, x' ∈ $\mathbb{R}^p$, p < d

e.g., projecting all features of a fruit to 2 dimensions for visualization



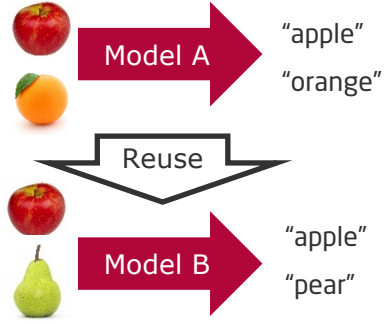
Semi-Supervised Learning (Some labels for training)

Anomaly / novelty detection

trained only on "normal" samples

e.g. x ∈ Apples, y ∈ { ☺ , ☹ }

Transfer Learning



Reinforcement Learning

Evaluation Exercise I

Data Management for Digital Health, Winter 2023

12

https://en.wikipedia.org/wiki/Apple,https://cdn4.vectorstock.com/i/1000x1000/16/58/robot-arm-line-icon-sign-on-vector-17841658.jpg

But...

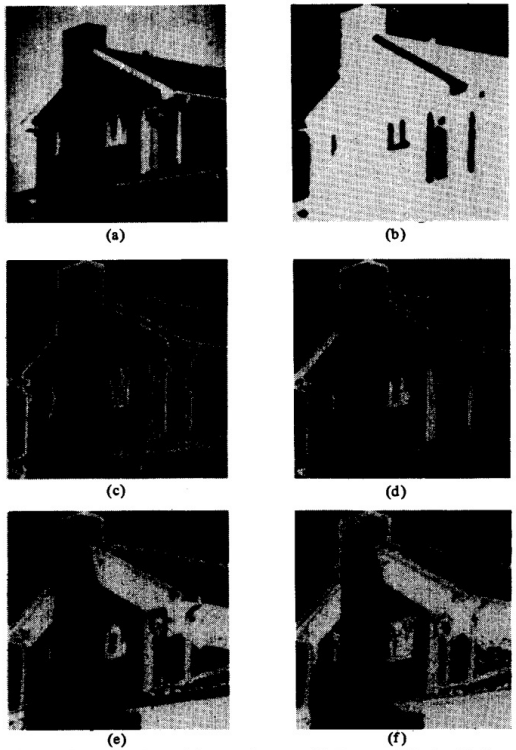


Fig. 17. Segmentation of house picture. (a) House original. (b) Two regions (best number of regions). (c) Three regions. (d) Four regions. (e) Five regions. (f) Six regions.

PROCEEDINGS OF THE IEEE, VOL. 67, NO. 5, MAY 1979

Image Segmentation by Clustering

GUY B. COLEMAN, MEMBER, IEEE, AND HARRY C. ANDREWS, SENIOR MEMBER, IEEE

Q13: What can you say about the different trained models? (Jupyter Notebook)

Hasso Plattner Institut

- ✓ Both standardization and polynomial features improve model performance in terms of area under the curve.
- ✗ The model with polynomial features (without standardization) decreases the number of false positives compared to the model trained on the raw input data.
- ✓ The model combining polynomial features with standardization decreases the number of false positives compared to the model trained on the raw input data.
- ✓ The model combining polynomial features with standardization has the highest precision.

Frequently missed

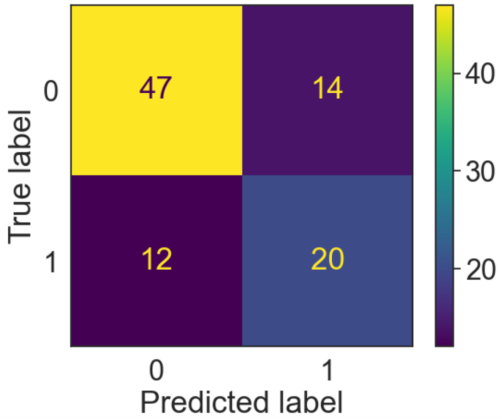
Frequent incorrect answer

Jupyter Notebook 1

Vanilla Logistic Regression Model

```
In [33]: do_eval(y_test, y_test_pred)
```

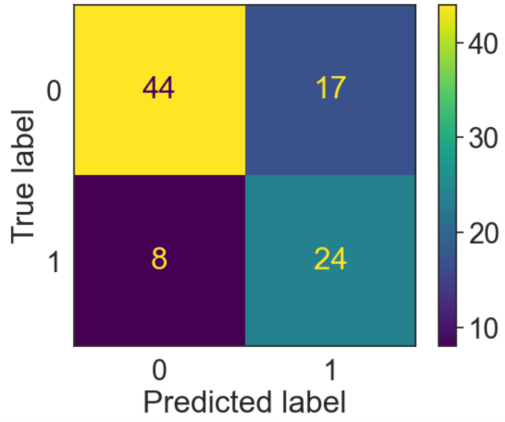
```
Accuracy: 0.72
Precision: 0.59
Recall: 0.62
AUC: 0.70
```



Logistic Regression Model + Polynomial Features

```
In [34]: do_eval(y_test, y_test_poly_pred)
```

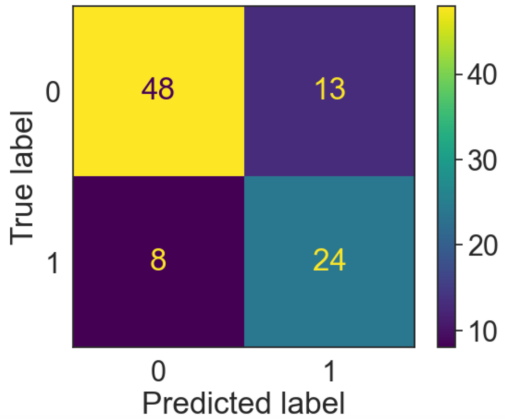
```
Accuracy: 0.73
Precision: 0.59
Recall: 0.75
AUC: 0.74
```



Logistic Regression Model + Polynomial Features + Standardization

```
In [36]: do_eval(y_test, y_test_poly_scaled_pred)
```

```
Accuracy: 0.77
Precision: 0.65
Recall: 0.75
AUC: 0.77
```



https://github.com/hpi-dhc/dm4dh-2023/blob/main/2_ML_Process.ipynb

Q20: Which of the following statements are correct for the terms Electronic Medical Record (EMR), Electronic Health Record (EHR) and Personal Health Record (PHR) as discussed in class.

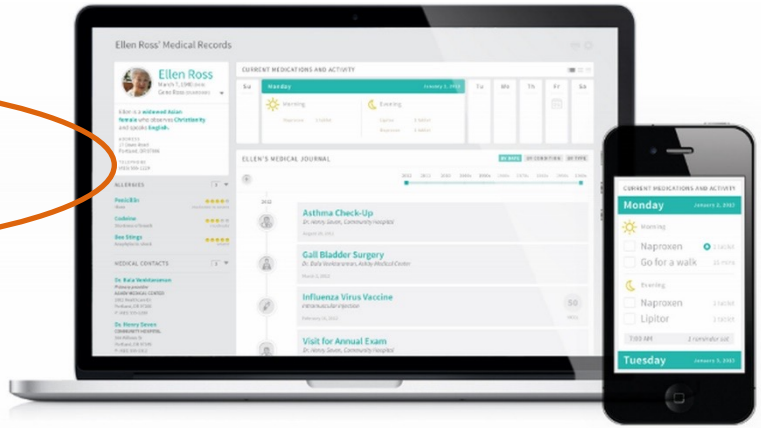
- ✗ All terms describe the same type of IT system.
- ✗ EMR focuses on medication data, e.g., from laboratory forms.
- ✓ EHR combines data from multiple sources, e.g., departments, facilities, to optimize healthcare processes.
- ✓ PHR focuses on the collection and combination of health information driven by the individual user and not necessary a medical professional

Frequently missed

Frequent incorrect answer

Electronic Medical Record

- Electronic Medical Record (EMR) :=
  Enabling technology, "just" electronic form of traditional (clinical) medical records
- Examples: medical history, diagnoses, medications, immunization dates, allergies
- Pro: Allows easy assessment of longitudinal data in digital form
- Cons:
  - Often limited to a single practice,
  - No process optimization,
  - No focus on compatibility & exchange with other systems



https://www.theverge.com/2013/1/28/3925734/is-nightingale-the-future-of-user-friendly-medical-records

EMR

Q22: Please select all statements that apply to the German health insurance card (eGK).

Hasso Plattner Institut

- ✗ You can maintain your personal details on your eGK, e.g., medication plan, emergency details, using public eGK card terminals provided in central places.
- ✗ Amongst others, on the eGK your personal patient history as well as your individual list of trusted medical doctors is stored.
- ✓ The eGK is required to setup access to your electronic patient record (ePA).
- ✓ The eGK is considered as the key component for data exchange in the German healthcare system

Frequently missed

Frequent incorrect answer

Vision: eGK Terminal made in Germany?

Evaluation Exercise I

Data Management for Digital Health, Winter 2023

German Electronic Health Insurance Card / Elektronische Gesundheitskarte (eGK) (cont'd)

- Mandatory for public health insurance
- Key component for data exchange in German healthcare system
- Stores personal (health) information, e.g. photo, name, birthdate, address, **insurance number**, etc. (see back of your card for stored details)
- (Future) applications: emergency details, electronic medication plan, e-prescription, etc.

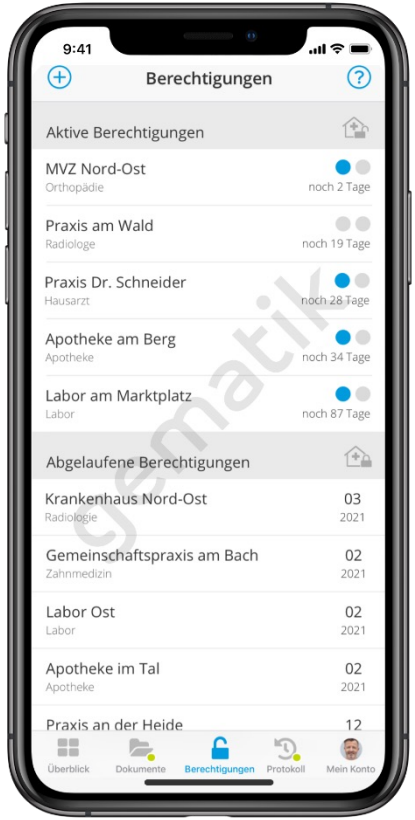
Source: Barmer

What is special about the German health insurance number?

German Electronic Patient Record: Elektronische Patientenakte (ePA)

Hasso Plattner Institut

- Germany 2023: Health insurances have to provide access to gematik's ePA.
- Lookup your health insurance and its corresponding app for accessing your ePA at: https://www.gematik.de/anwendungen/e-patientenakte/epa-app
- Download app, register app with your health insurance card, access your ePA (even after switching insurances).



Latest News from Digital Summit 2023

Q24: Please select all appropriate answers as discussed in class.

- ✗ A coding system can be used as replacement for a terminology or a classification.
- ✗ A terminology groups classes of similar terms from different languages together to support internationalization in software products.
- ✓ A classification is used to arrange selected terms, e.g., cold, rhinitis, according to an established concept to form a category or group, e.g., upper respiratory infection.
- ✓ Coding systems, terminologies, and classifications facilitate interoperability.

Frequently missed

Frequent incorrect answer

Interoperability:
Classifications, Terminologies, Coding Systems

- Classification := Arrangement of facts, terms, conditions according to a established criteria or concept to form a group or category.
  - International Classification of Diseases (ICD)
  - Anatomical Therapeutic Chemical (ATC)
- Terminology := Set of terms for a particular field not necessary arranged in groups
  - SNOMED CT: medical observations, findings, etc.
  - Unified Medical Language System (UMLS): maps vocabularies and terms of terminologies.
- Coding System := Unified definition of criteria and valid ranges
  - Unified Code for Units of Measure (UCUM): focus medicine and pharmacy
  - Logical Observation Identifiers Names and Codes (LOINC): encoding of sample details, procedures, lab results