Maschinelles Lernen

Introduction Data

Maschinelles Lernen

- Credits for content and many slides:
  - Introduction to Data Mining by Tan, Steinbach and Kumar, Addison Wesley (many slides & 3 chapters on line)
  - Data Mining by Han and Kamber, Morgan Kaufmann (many slides on line)

How does it work?

How does it work?

How does it work?

Kunden, die diesen Artikel gekauft haben, kauften auch

Data Science for Business: What you need to know about data mining and…
› Foster Provost
★★★★½ 8
Taschenbuch
EUR 27,69 ✓prime

Knowledge Discovery in Databases: Techniken Und Anwendungen
› Martin Ester
★★★★★ 1
Taschenbuch
EUR 49,99 ✓prime
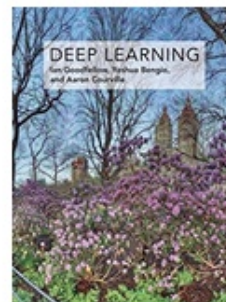
Deep Learning (Adaptive Computation and Machine Learning)
› Ian Goodfellow
★★★½ 18
Gebundene Ausgabe
EUR 39,49 ✓prime

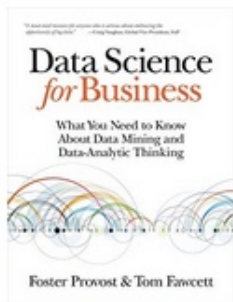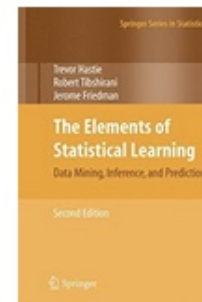**The Elements of Statistical Learning: Data Mining, Inference, and…**
Trevor Hastie
★★★½ 10
Gebundene Ausgabe
EUR 37,05

Hands-On Machine Learning with Scikit-Learn and TensorFlow:…
› Aurélien Géron
★★★★★ 18
Taschenbuch
EUR 33,72 ✓prime

Process Model

Cross Industry Standard Process for Data Mining
http://www.crisp-dm.org/Process/index.htm

What is Data?

- Collection of data objects and their attributes

- An attribute is a property or characteristic of an object
  - Examples: eye color of a person, temperature, etc.
  - Attribute is also known as variable, field, characteristic, or feature
- A collection of attributes describe an object
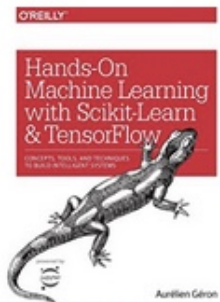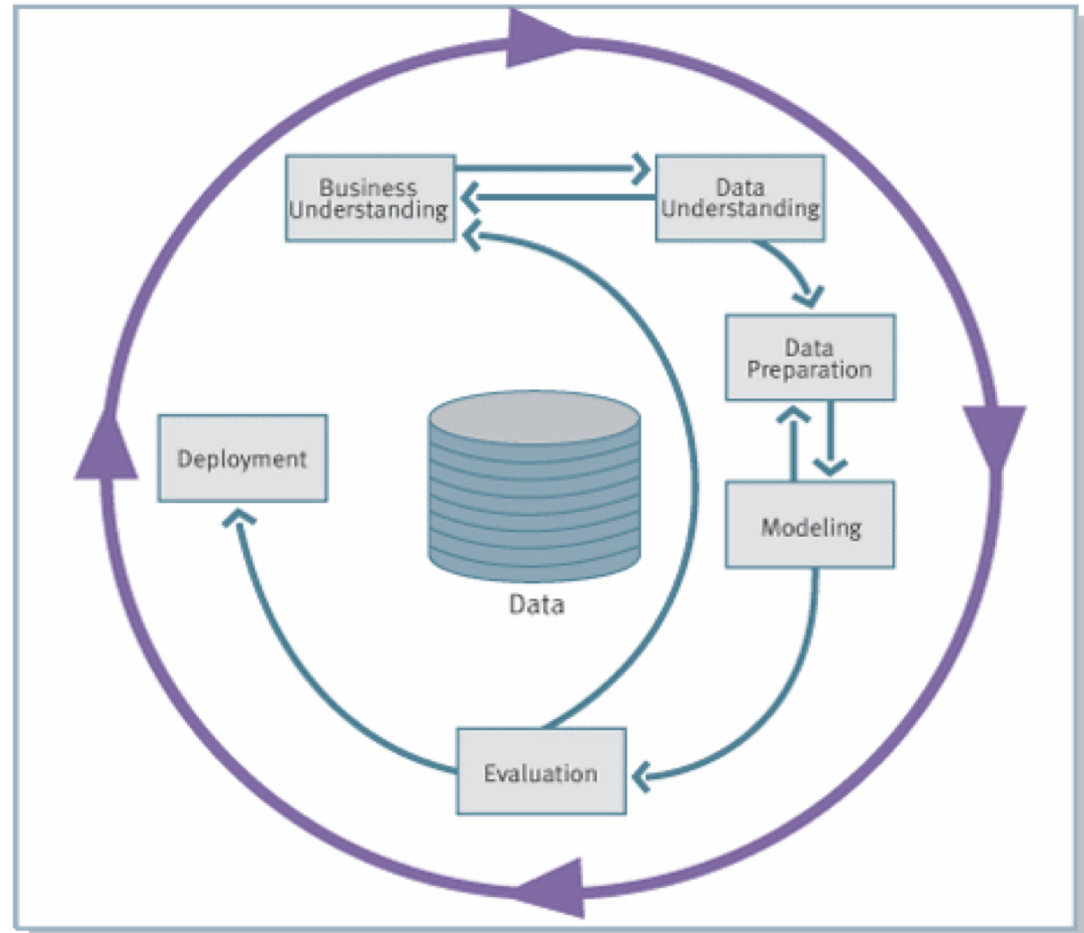 - Object is also known as record, point, case, sample, entity, or instance

Attributes

Objects

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

The Iris Data-Set

● Iris setosa

The Iris Data-Set

@RELATION iris
@ATTRIBUTE sepallength   REAL
@ATTRIBUTE sepalwidth   REAL
@ATTRIBUTE petallength   REAL
@ATTRIBUTE petalwidth   REAL
@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}
@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
6.3,3.3,4.7,1.6,Iris-versicolor
4.9,2.4,3.3,1.0,Iris-versicolor
6.6,2.9,4.6,1.3,Iris-versicolor
7.7,2.8,6.7,2.0,Iris-virginica
6.3,2.7,4.9,1.8,Iris-virginica

Discrete and Continuous Attributes

- Discrete Attribute
  - Has only a finite or countably infinite set of values
  - Examples: zip codes, counts, or the set of words in a collection of documents
  - Can be ordered: small, medium, large
  - Note: binary attributes can be a special case of discrete attributes like yes/no, male/female.

- Continuous Attribute
  - Has real numbers as attribute values
  - Examples: temperature, height, or weight
  - Practically, real values can only be measured and represented using a finite number of digits
  - Note: binary attributes can be a special case of continuous attributes 0/1.

Data Preprocessing – Übung 1

- Data cleaning

  Right values

- Data transformation

  In the right form

- Data integration

  All together