

Machine Learning für Zeitreihen: Überblick, Klassifikation

Session 6 (Dienstag 11:00 – 12:30)



Machine Learning für Zeitreihen: Überblick, Vorbereitung und Vorhersagen

- Besonderheiten Machine Learning für Zeitreihen
- Random Forest Klassifikator, Bsp: Belegungserkennung anhand von Raumklimadaten

KI, Machine Learning und Deep Learning für Zeitreihen

- **Machine Learning und Deep Learning** revolutionieren Umgang mit Daten
- Was ist mit Zeitreihen?
- Die meisten Verfahren sind **nicht speziell für Zeitreihen** (Ausnahmen später)
- Sie können aber für den Umgang mit Zeitreihen (**sehr erfolgreich**) genutzt werden
- Entscheidend ist:



KI, Machine Learning und Deep Learning für Zeitreihen

Artificial Intelligence (AI)



Component technologies

Expert systems

Robotics

Human interfaces

⋮

Machine learning



Algorithms

SVM

Random Forest

K-means

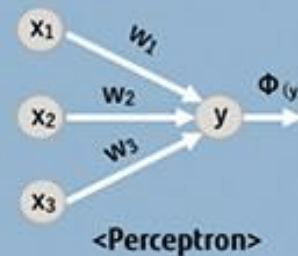
⋮

Reproduction of human intellectual activities

Learning of rules and patterns hidden in data

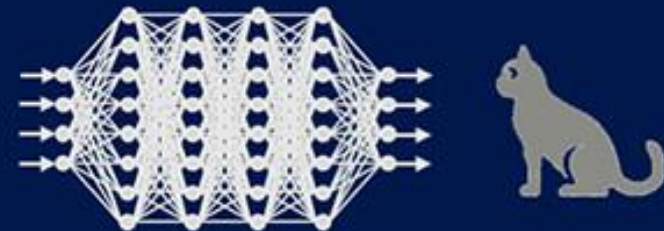
Neural networks

Learning model that emulates brain activities



Modeling of neural signaling

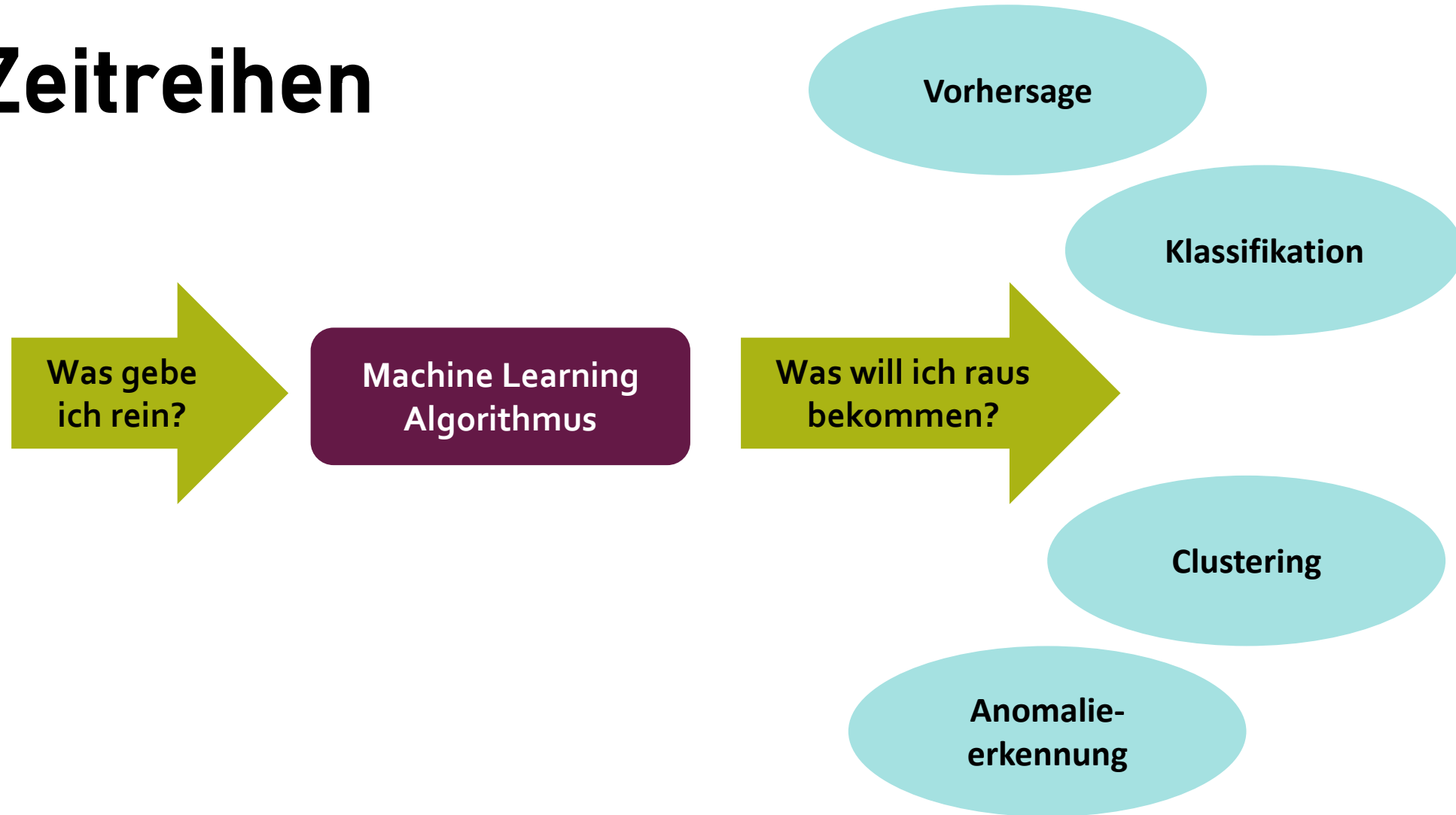
Deep Learning
Learning using multi-layer neural networks



ML für Zeitreihen



ML für Zeitreihen



Rollendes Zeitfenster

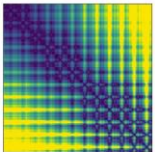
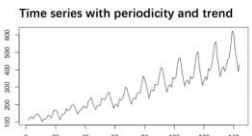
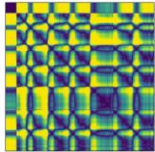
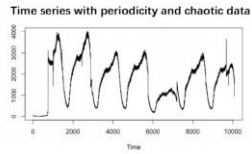
Daten ungeordnet

11786.5 23.6

11220.0 24.0

10444.1 24.3

... ...



Zeitreihe als Bild

“Feature Extraction” Merkmale

Was gebe
ich rein?

Machine Learning
Algorithmus

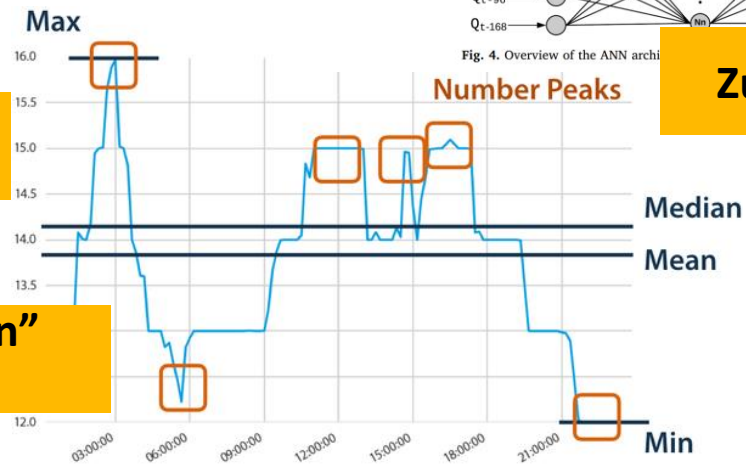
Was will ich raus
bekommen?

Vorhersage

Klassifikation

Clustering

Anomalie-
erkennung



calculate
features predict

calculate
features predict

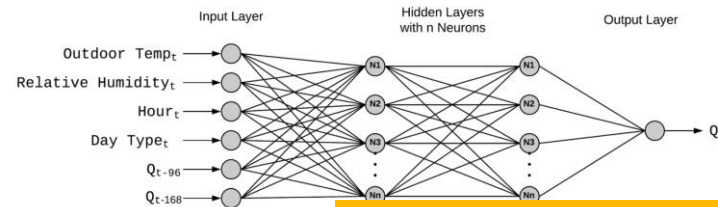


Fig. 4. Overview of the ANN archi

Zusätzliche Merkmale

Beispiel: Belegungserkennung eines Büroraums aus Temperatur-, Feuchte-, Licht- und CO2-Messungen

Daten ungeordnet

11786.5	23.6
---------	------

11220.0	24.0
---------	------

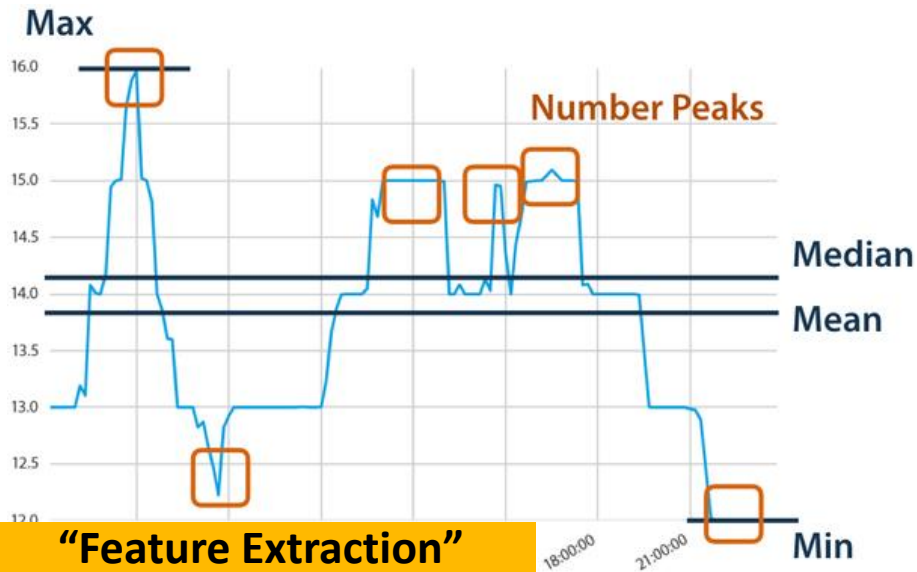
10444.1	24.3
---------	------

...	...
-----	-----

Sekunden
Entfernung von
Tagesmitte

Werktag (1/0)

Arbeitsstunden
8:00 – 20:00 Uhr
(1/0)



“Feature Extraction”
Merkmale

Zusätzliche Merkmale

Temperatur

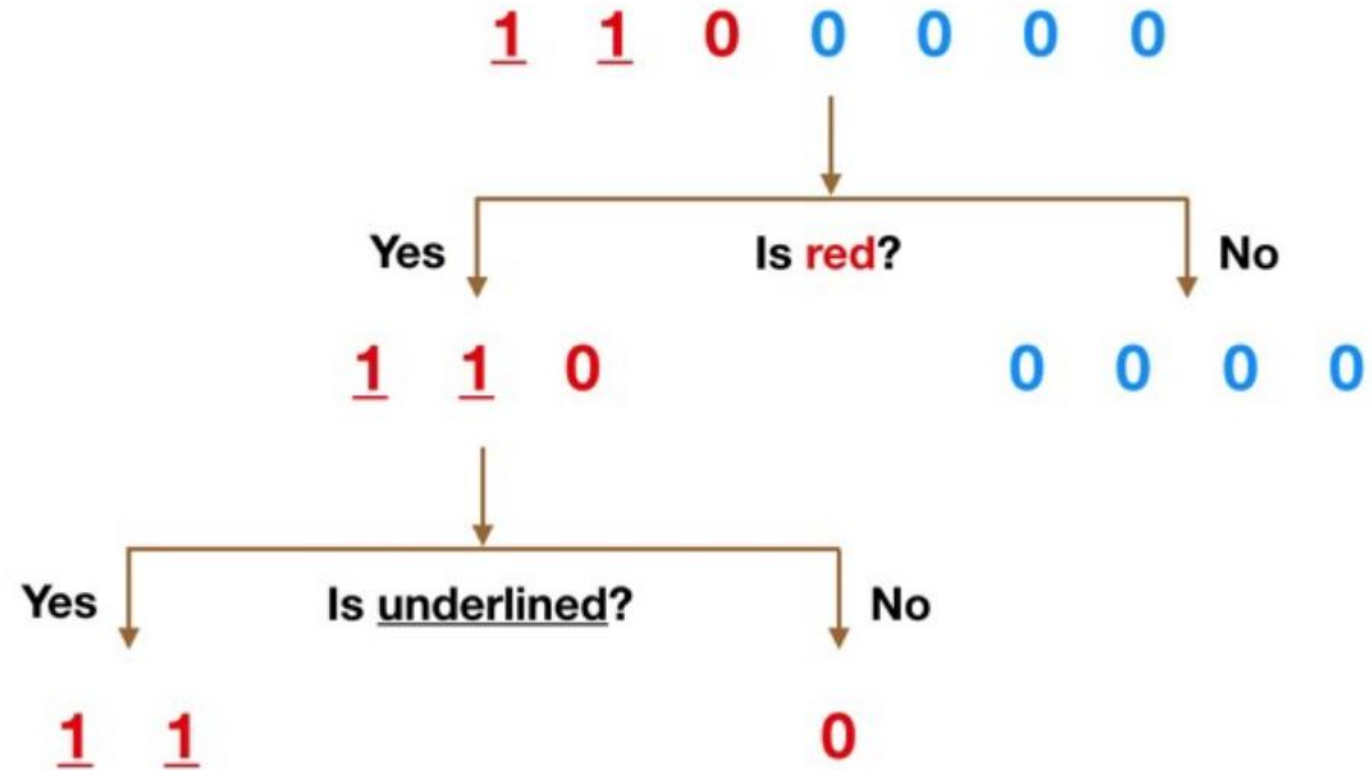
Luftfeuchte

Licht

CO²

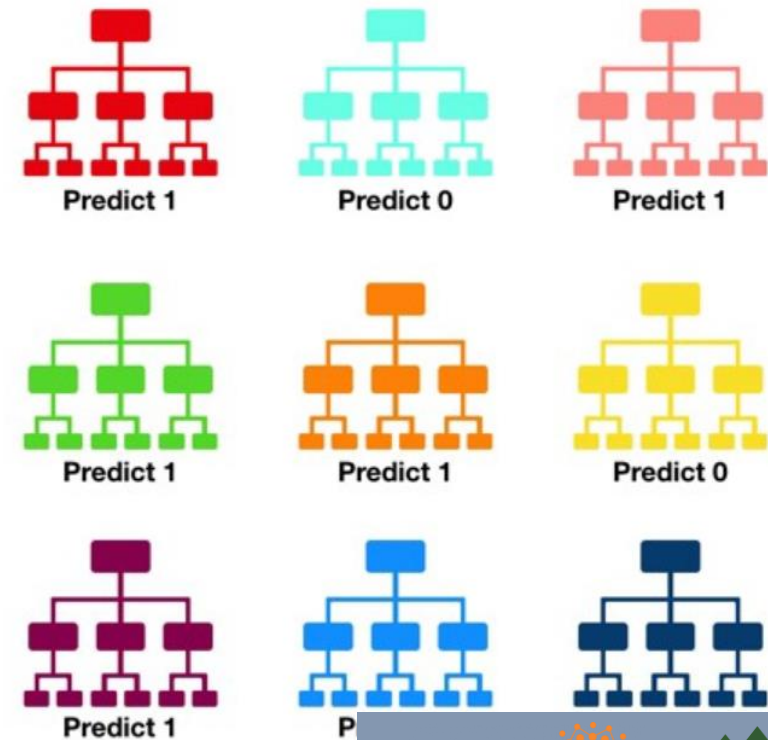
Belegung

Decision Trees



Random Forests

- Der Random Forest besteht, wie der Name schon sagt, aus einer großen Anzahl von einzelnen Entscheidungsbäumen, die als Ensemble arbeiten.
- Jeder einzelne Baum im Random Forest spuckt eine Klassenvorhersage aus und die Klasse mit den meisten Stimmen wird zur Vorhersage unseres Modells



Gini Impurity

- Maß dafür, wie oft ein zufällig ausgewähltes Element aus der Menge falsch beschriftet würde, wenn es entsprechend der Verteilung der Beschriftungen in der Teilmenge zufällig beschriftet würde.
- Die Gini Impurity lässt sich berechnen, indem man die Wahrscheinlichkeit, dass p_i gewählt wurde, mit der Wahrscheinlichkeit
$$\sum_{k \neq i} p_k = 1 - p_i$$
eines Fehlers bei der Einstufung dieses Elements multipliziert.
- Sie erreicht ihr Minimum (Null), wenn alle Fälle des Knotens in eine einzige Zielkategorie fallen.

CODING

Belegungserkennung

- Klassifikation

Verbrauchergruppen

- Clustering

Ab ins Jupyter
Notebook

