Conformal Prediction

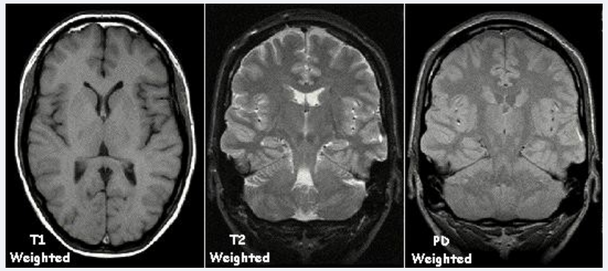
Unsicherheit von KI-Modellen quantifizieren

Dr. Tobias Quadfasel | Data Scientist @ Ailio

Warum Unsicherheit quantifizieren?

ML-Modelle:
Lernen Muster aus Daten, um **Vorhersagen** zu machen
→ Brauchen Informationen, wie **sicher** Modell ist

KI-Anwendungen in kritischen Bereichen:



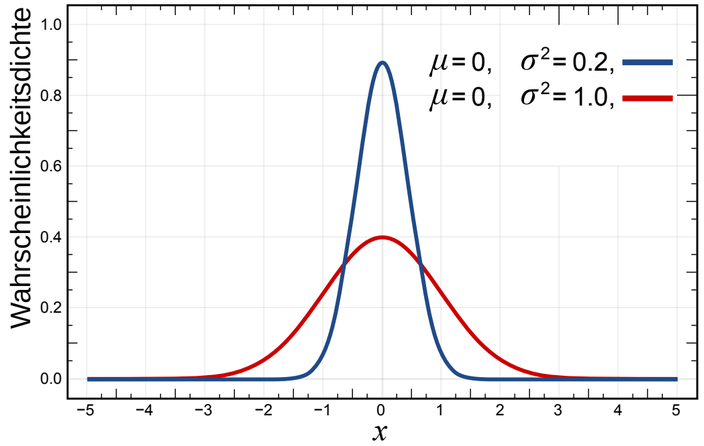
Medizinische Diagnostik

Geschäftskritische Entscheidungen

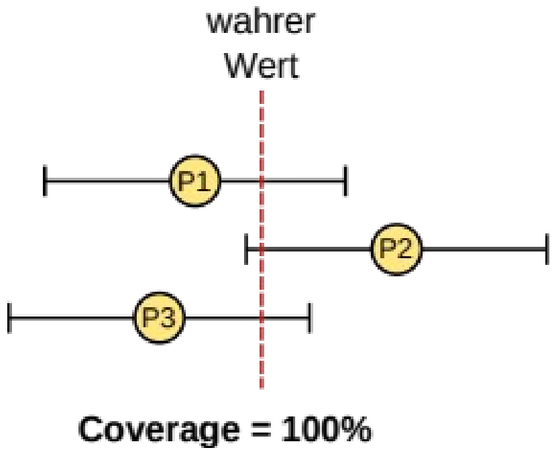
Anomalie-Erkennung/ Fraud Detection

Unsicherheit: Was ist das?

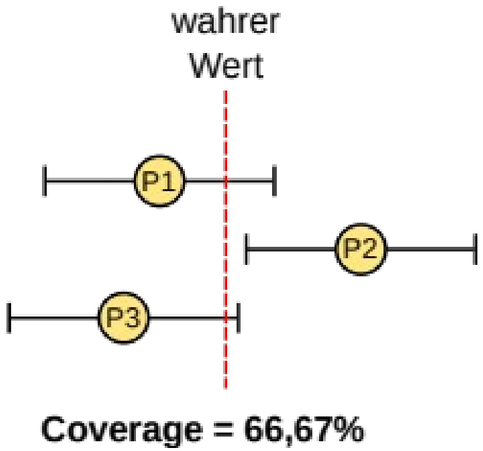
- Statistische Beschreibung über **Varianz/Standardabweichung**



- Wichtiges Kriterium: **Coverage** (Überdeckungswahrscheinlichkeit)



Coverage = 100%



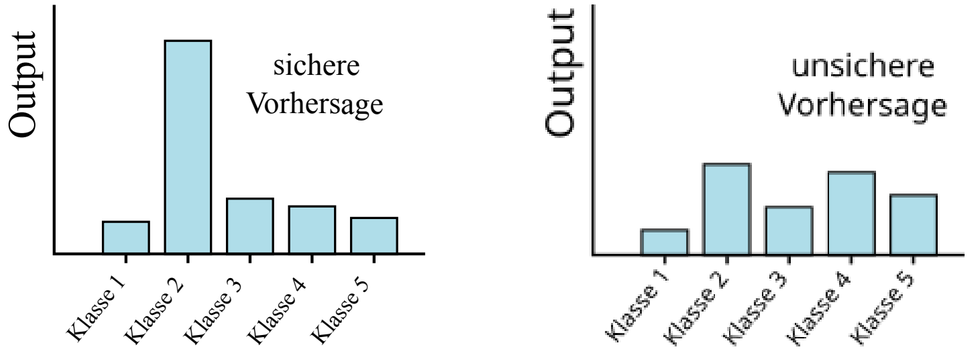
Coverage = 66,67%

Quellen von Unsicherheit in ML Modellen

- Auswahl des Trainings-Datensatzes
- Trainings-Algorithmus
- Hyperparameter Optimierung / Model Selection
- Fehlende Werte / Noise

Unsicherheit in ML Modellen: Methoden

- **Klassifikation:** Modell-Output



→ fehlende Kalibrierung

- **Bayes-Methoden:** Posterior-Verteilung

→ Abhängigkeit von Prior + Datenannahmen

- **Naive Methoden:**

→ **Random forest**: Varianz der Trees

→ **Bootstrapping**: erneutes Fitten des Modells mit anderen Samples

Problem: Keine mathematischen Garantien für die **Coverage**

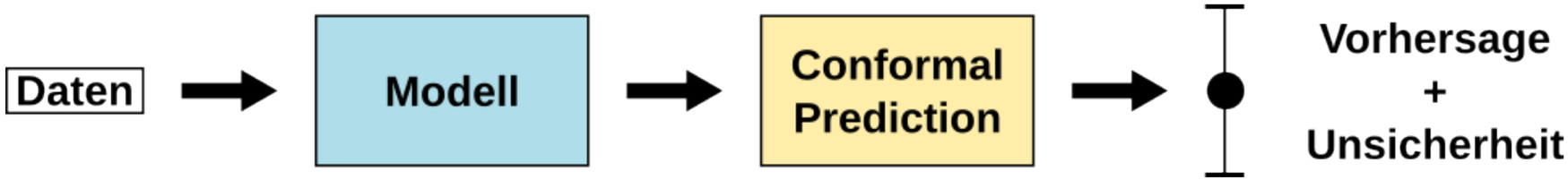
Conformal Prediction: Vorteile

- Mathematisch garantierte Coverage
- Funktioniert für nahezu alle ML Use-Cases
  - Klassifikation
  - Regression
  - Time Series Forecasting
  - Anomaly Detection
  - ...
- Funktioniert modell-unabhängig
  - Ensemble Models, neuronale Netze, Clustering-Algorithmen, ...
- Minimaler zusätzlicher Rechenaufwand
- Software-Packages vorhanden (z.B. für python)

Conformal Prediction: Grundlagen

Klassisches Machine Learning:

Conformal Prediction:



Beispiel: Klassifikation

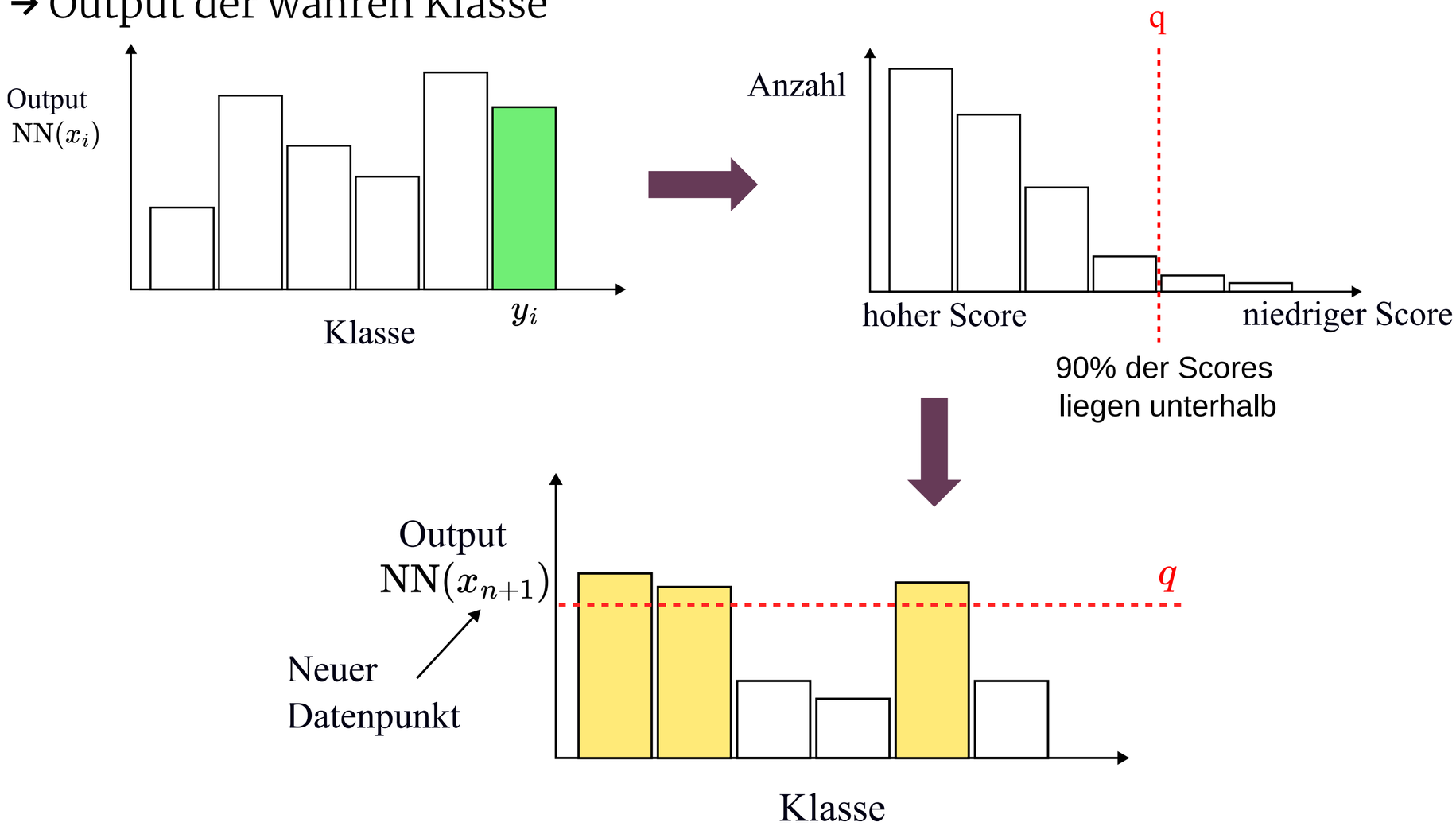
Anstatt **Punktvorhersage** liefert Conformal Prediction eine **Vorhersage-Menge (prediction set)**

Ailio

CP Beispiel: Klassifikation

Berechne für alle Datenpunkte (x_i, y_i) den uncertainty score E_i

→ Output der wahren Klasse



Ailio

Conformal Prediction mit Python

Wichtigste Python-Bibliothek: Mapie

- Einfache Berechnung von konformalen Vorhersage-Intervallen bzw. -Mengen
- CP-Algorithmen für Regression, Klassifikation und Zeitreihen-Analyse
- Bietet Schnittstelle mit allen gängigen ML-Bibliotheken (scikit-learn, pytorch, tensorflow,...)
- Scikit-learn kompatibler Wrapper

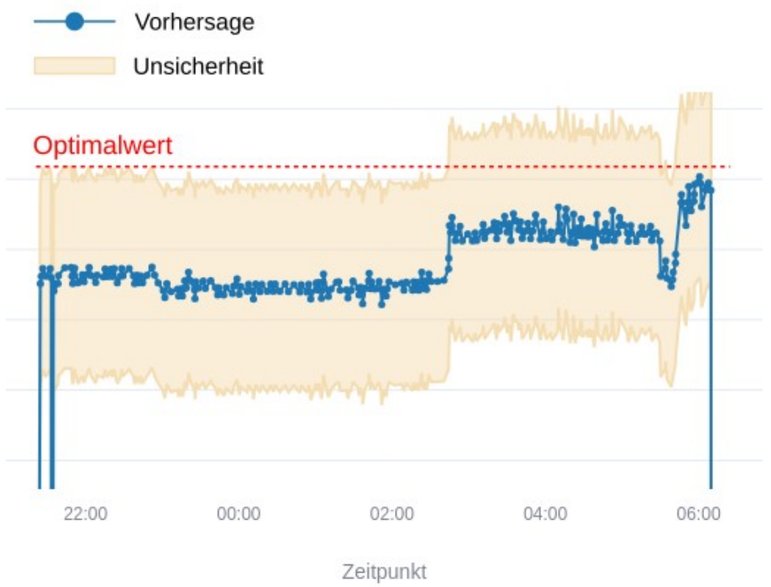

```python
from mapie.classification import MapieClassifier

# model is pre-fitted ML model
cp = MapieClassifier(estimator=model, cv="prefit", method="score")
cp.fit(X_calib, y_calib)

# Evaluate Models on new Datapoint X_new
# y_pred -> predictions from original model
# y_unc -> conformal prediction sets
y_pred, y_unc = cp.predict(X_new, alpha=0.05)
```

Ailio

CP Use-Case: Prozess-optimierung



- Hergestelltes Produkt muss bestimmte Kriterien erfüllen
- → Vorhersage des Merkmals mit ML Modell
- Maschinen sollen möglichst nah am optimalen Wert gefahren werden
- CP-Intervalle: Bei längerem Unterschreiten des Optimalwerts, Meldung an Maschinenführer
- Zusätzliche Insights durch Unsicherheiten des Modells

Conformal Prediction: Use-Cases

Medizinische Diagnostik:

- CP in der **Brustkrebs-Erkennung**: Link
- Erkennung von **Schlaganfall-Risiko** Anhand von Ultraschallbildern mit ML + CP: Link

Medikamentenentwicklung:

- Vorhersage von Eigenschaften **chemischer Verbindungen** mit ML + CP: Link

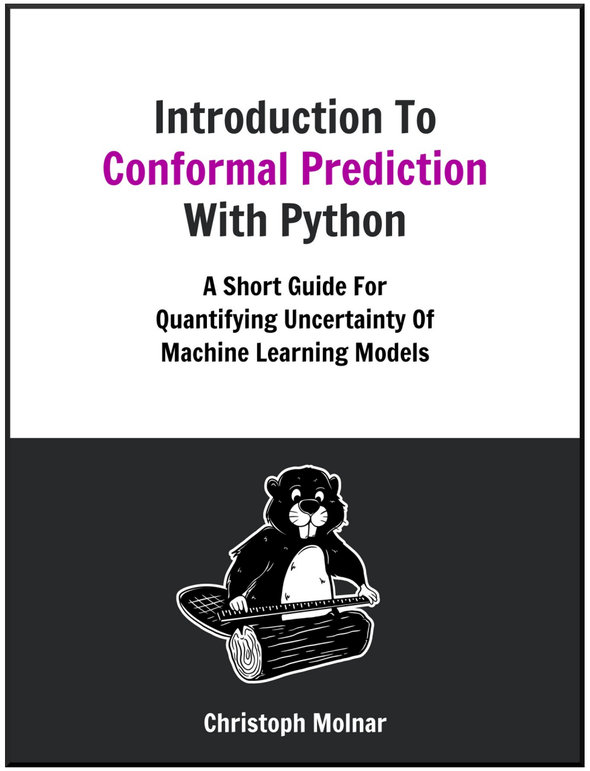
Verkehr & Infrastruktur:

- Bestimmung der Unsicherheit von **Verkehrsprognosen** mit CP: Link
- Robuste **Gaspreis-Vorhersage** mit CP: Link

LLMs:

- **“Robots that ask for help”**: LLM Agenten lernen mit CP, wann sie etwas nicht wissen und sich Hilfe holen müssen (DeepMind): Link
- “Prompt Risk Control”: Minimierung schädlicher Prompts für **Responsible LLM** deployment mit CP: Link

Weiterführende Literatur & Infos



- Valery Manokhin: “Awesome Conformal Prediction” repo auf Github: Link
- Einführungs-Paper: Link
- Youtube Channel von Anastasios Angelopoulos: Link

Alternative zu MAPIE von AWSlabs:

Fortuna (Link)

→ Einfache SageMaker Integration

Ailio