

# **Sitzung 3 – Version 2**

# Inhalt

---



## Modellvalidierung

4.1 Rückblick auf die letzte Veranstaltung

4.2 Gütemaße

4.3 Varianz und Bias



## Projektabschluss

5.1 Antwort auf das Stahlprojekt

5.2 Ausblick

# 4 Modellvalidierung

---

4.1 Rückblick auf die letzte Sitzung

4.2 Gütemaße

4.3 Varianz und Bias

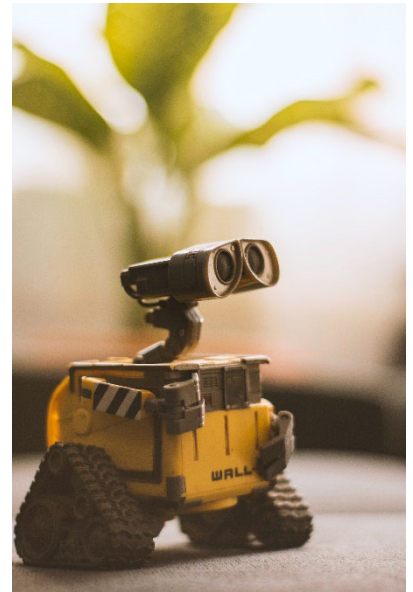


Foto von Lenin Estrada von Pexels

# Lernziele

---

1. Sie nutzen die Visualisierung der Testphase eines überwachten Lernverfahrens, um erste Aussagen über die Qualität des entstandenen Modells zu treffen.
  2. Sie stellen die Ergebnisse der Testphase eines binären Klassifizierungsverfahrens mit Hilfe der Confusion-Matrix und des statischen Einheitsquadrates dar.
  3. Sie interpretieren das Gütemaß Accuracy als prozentuale Trefferquote eines Klassifizierungsverfahrens und visualisieren den Wert mit Hilfe des Einheitsquadrates.
  4. Sie visualisieren Precision und Recall sowie deren Zusammenhang im Einheitsquadrat.
-

# Lernziele

---

- 5. Sie erläutern die Interpretation und Aussagekraft von Accuracy, Precision, Recall und  $F_1$ -Score unter Verwendung typischer Beispiele.
- 6. Sie trennen in Python einen Datensatz in Testdatensatz und Trainingsdatensatz und bestimmen gängige Gütemaße entstandener Modelle bezüglich der Testdaten.
- 7. Sie erkennen typische Probleme wie Varianz, Bias, Overfitting und Underfitting und zugehörige Einflussfaktoren beim Fitten von Modellen mit überwachten Lernverfahren.

# 4.1 Rückblick auf die letzte Sitzung

---

# Stahlprojekt: Ausgangssituation

---

Sie arbeiten für eine Firma, die Bauteile aus Stahl herstellt und vertreibt.

Die Entwicklungsabteilung Ihrer Firma hat einen neuen Stahl ausgewählt, welcher unter hoher Belastung besonders langlebig sein soll. In einem bestehenden Produktionsprozess für Bauteile könnte dieser Stahl den bisher verwendeten Stahl ersetzen.

Es soll nun ein ML-Modell entwickelt werden, um aus der Anzahl der Risse & der Breite des größten Risses die Stahlqualität zu beurteilen.

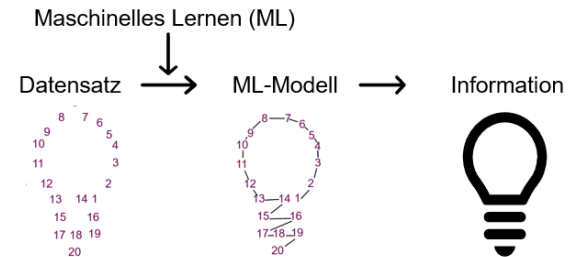


Foto von Martinelle auf Pixabay

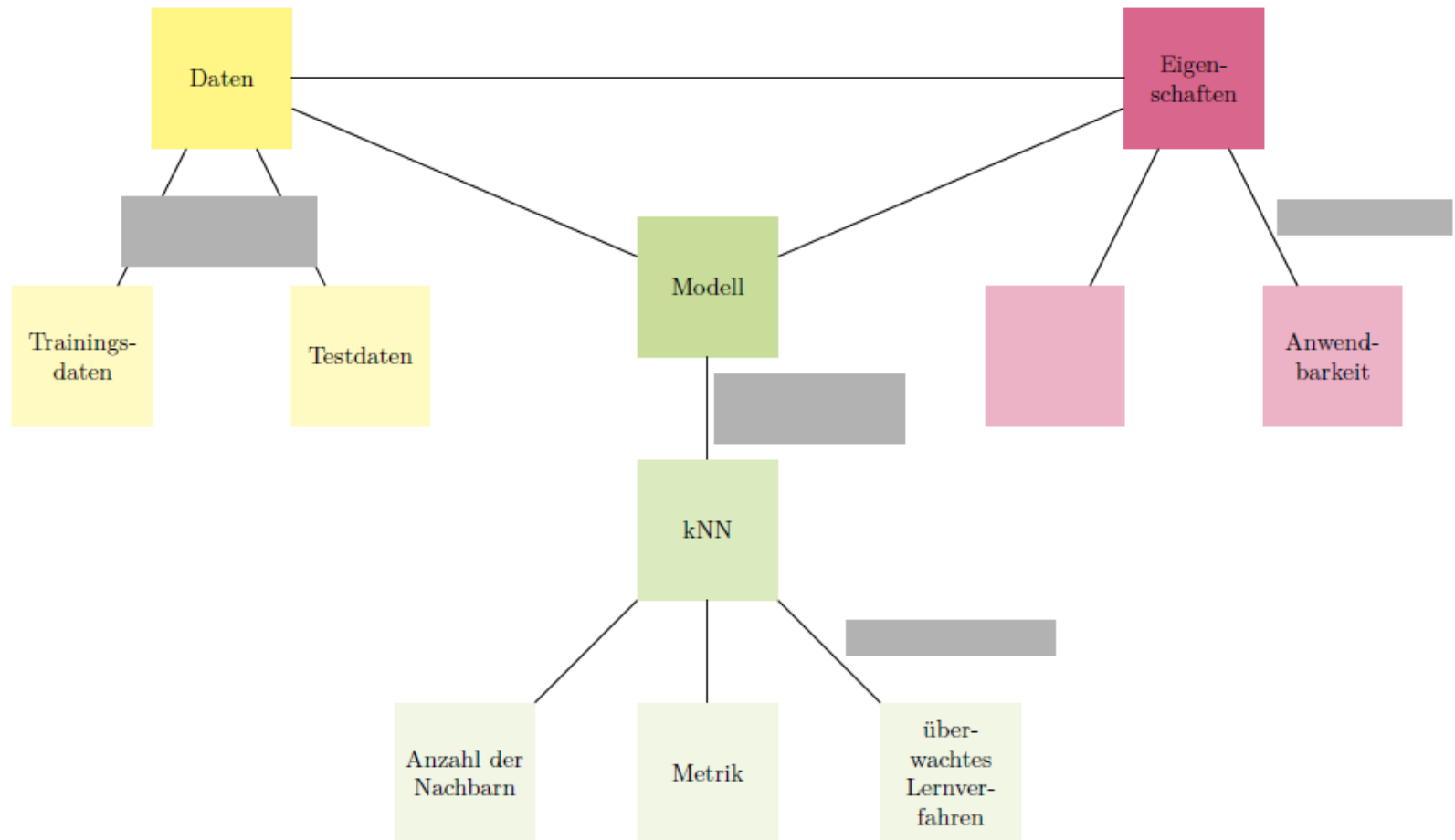
# Rückblick auf die *Aufgabe zum Abschluss*

Bearbeiten Sie mit Rückblick auf die heutige Veranstaltung die drei folgenden Arbeitsaufträge:

1. Beschreiben Sie in 1-2 Sätzen, was ein ML-Modell im Kontext überwachtes Lernen ist.
2. Notieren Sie 3 Aspekte, an denen Sie die Qualität eines Modells beurteilen würden.
3. Beschriften Sie die nachstehende Grafik so, dass der Zusammenhang zwischen den einzelnen Begriffen durch die Grafik dargestellt wird. Erweitern Sie die Grafik um weitere in diesem Kontext relevante Begriffe und Verbindungen.







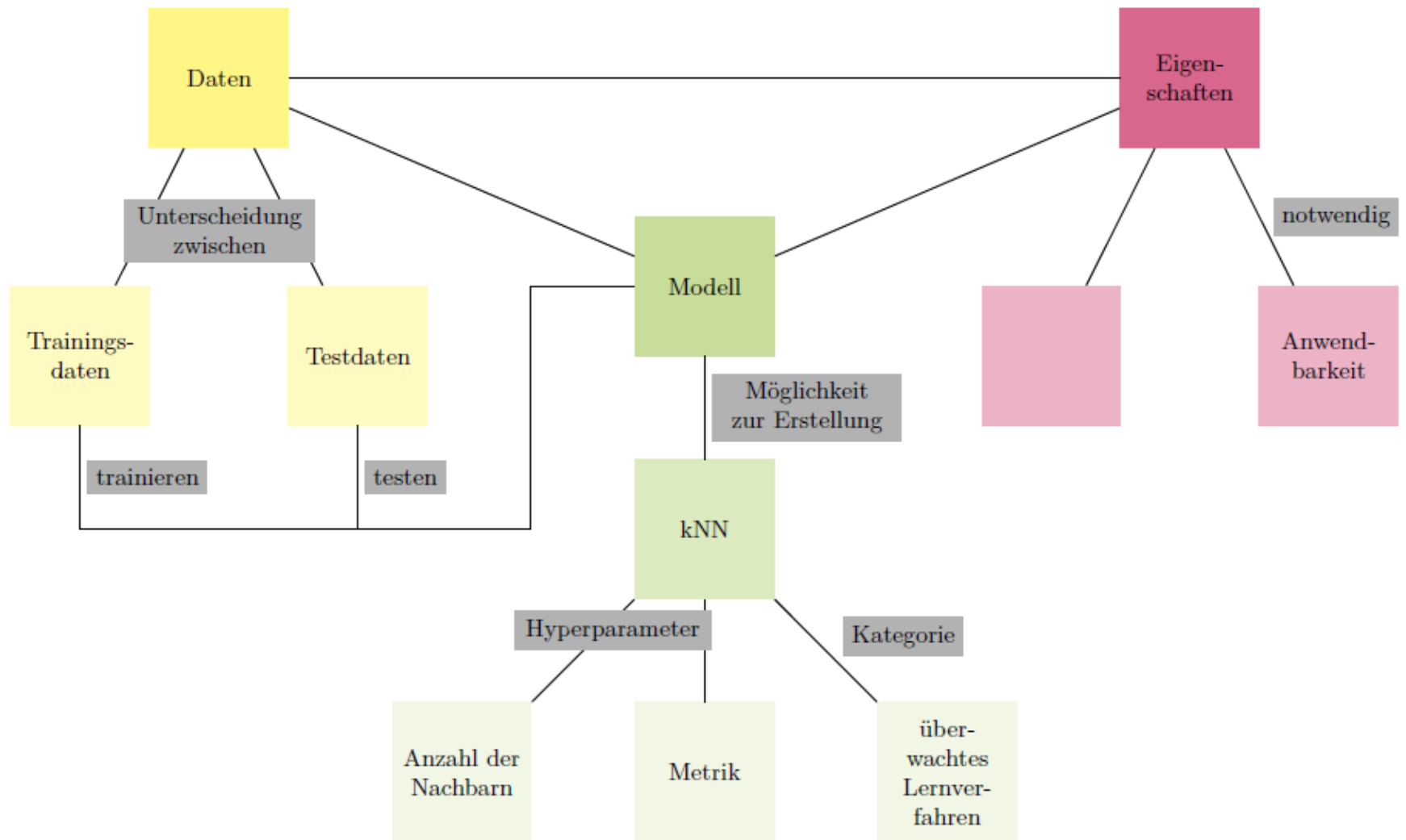
# Aufgabe zum Abschluss – Lösung

---

1. Beschreiben Sie in 1-2 Sätzen, was ein ML-Modell im Kontext überwachtes Lernen ist.

*Ein ML-Modell im Kontext überwachtes Lernen ist das Ergebnis der Anwendung eines ML Algorithmus auf einen Datensatz. Hierbei wird eine Zuordnung vom Feature zum Label entwickelt.*

2. Notieren Sie 3 Aspekte, an denen Sie die Qualität eines Modells beurteilen würden.
    - *Das Modell zeigt eine zu den Trainingsdaten passende Zuordnung von Features zum Label.*
    - *Die Zuordnung des Modells ist einfach zu verstehen.*
    - *Das Modell ist für die zugrundeliegende Fragestellung sinnvoll anwendbar.*
-



# Stahlprojekt:

## Modellqualität

---

Die Abbildung auf der nächsten Folie zeigt ein mit der kNN-Klassifikation gefittetes Modell ( $k=1$  und euklidische Metrik) mit den zugehörigen Testdaten.

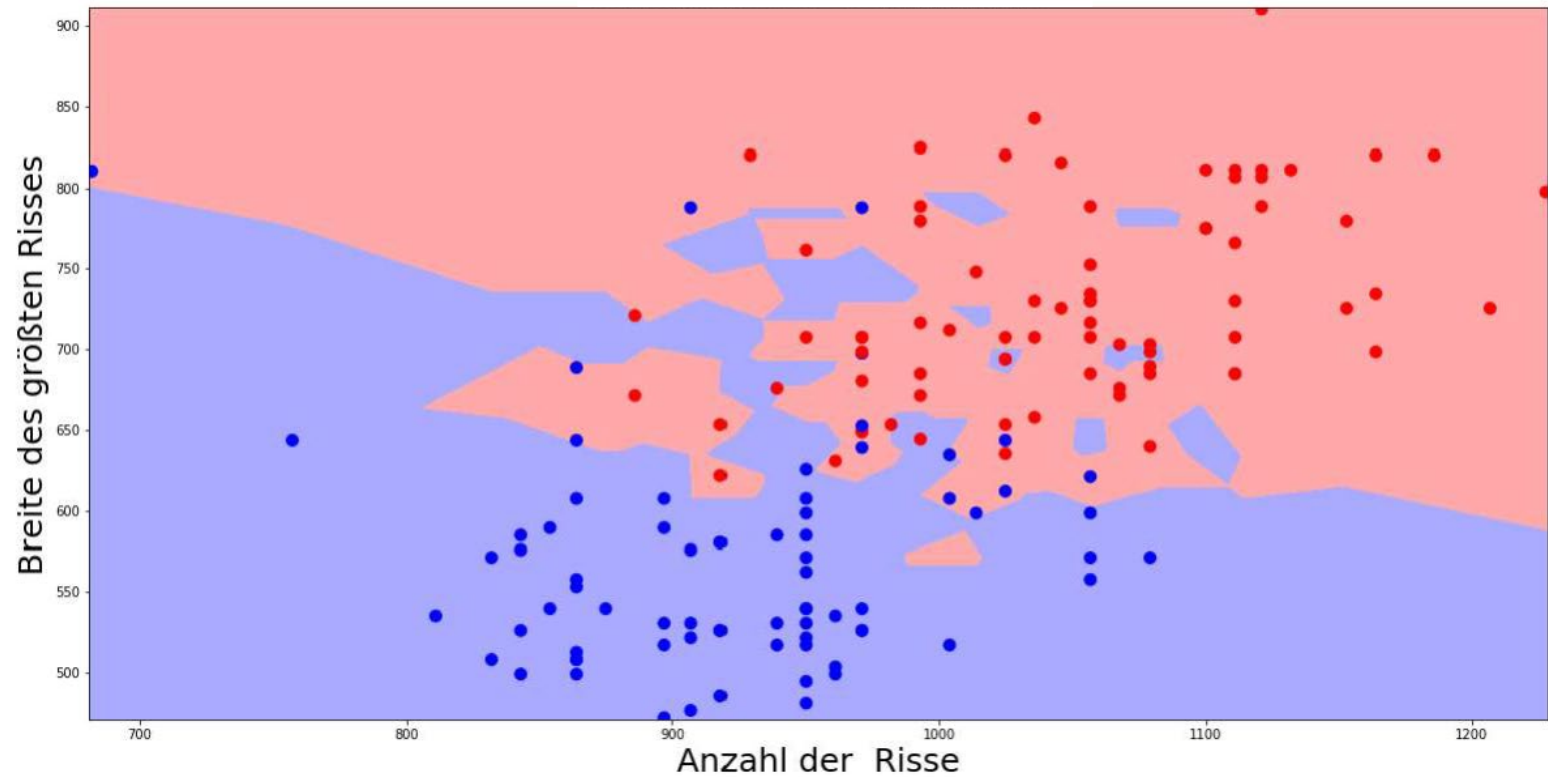
### Arbeitsauftrag:

Diskutieren und notieren Sie die Antworten auf die folgenden Fragen. Nutzen Sie hierfür Mural (Link folgt in Veranstaltung).

- Würden Sie aus der Grafik folgern, dass es sich um ein „gutes“ Modell handelt? Benennen Sie unterschiedliche Gründe.
- Formulieren Sie begründet allgemeingültige Kriterien für die Beurteilung eines Modells.

# Stahlprojekt: Visualisierung

---



# Stahlprojekt: Modellqualität

---

## Frage 1:

- ☹ Es gibt sehr viele Punkte, deren Farbe nicht zum zugehörigen Hintergrund passt, das lässt nicht auf ein gutes Modell schließen.
- ☹ Im Anwendungskontext wirkt es fragwürdig, dass sich keine homogene Flächen bilden, sondern Inseln, das lässt nicht auf ein gutes Modell schließen.

## Frage 2:

- ☹ Eine Klasse wird gut erkannt (wenig falsche Klassifikationen im Vergleich zu den Testdaten bei einer Klasse)
  - ☹ Alle Klassen werden gut erkannt (wenig falsche Klassifikationen im Vergleich zu den Testdaten bei allen Klassen)
  - ☹ Das Modell ist einfach (darstellbar / formulierbar)
  - ☹ Das Modell passt zum Anwendungskontext
-

## 4.2 Gütemaße

---

# Die Confusion-Matrix

In der Testphase eines Modells mit zwei Klassen werden vier Arten von Beispielen unterschieden:

		Tatsächliche Klasse	
		Klasse 1	Klasse 2
Vorhergesagte Klasse	Klasse 1	$a_{11}$	$a_{12}$
	Klasse 2	$a_{21}$	$a_{22}$

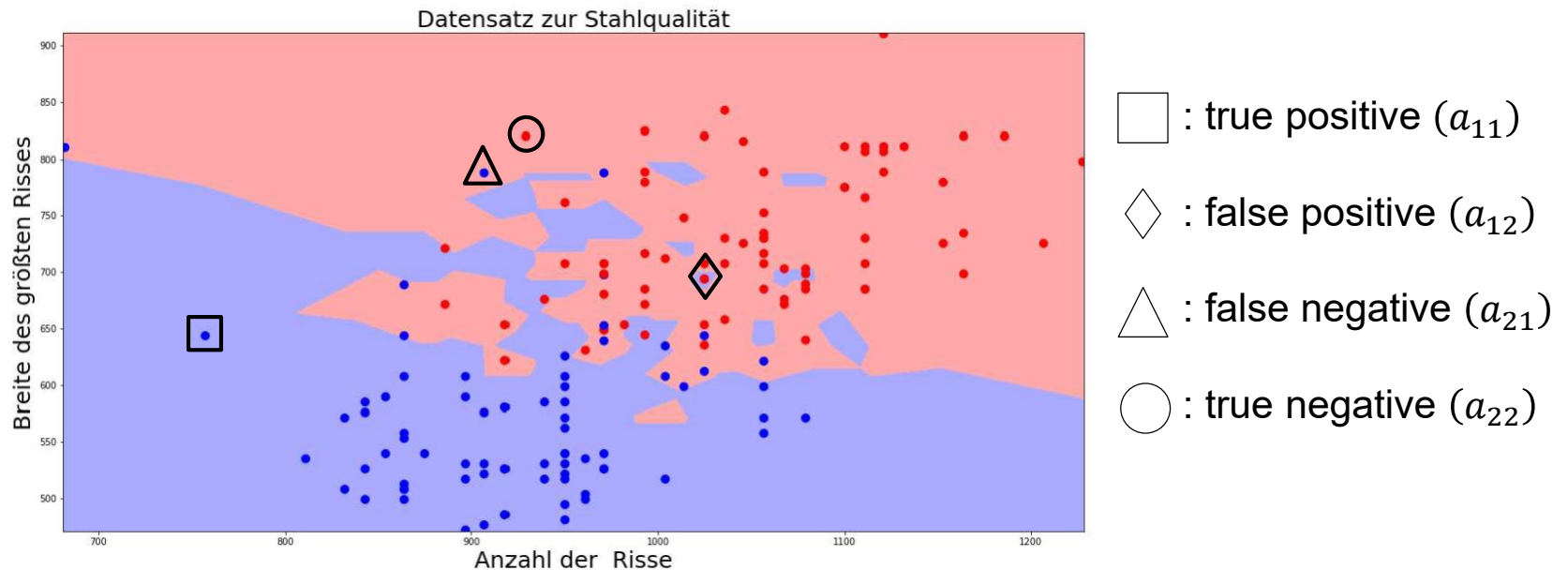
Alternativ:

		Tatsächliche Klasse	
		Kl. 1 (positive)	Klasse 2 (negative)
Vorhergesagte Klasse	Kl. 1 (positive)	$tp$ (true positive)	$fp$ (false positive)
	Kl. 2 (negative)	$fn$ (false negative)	$tn$ (true negative)



# Die Einträge der Confusion-Matrix

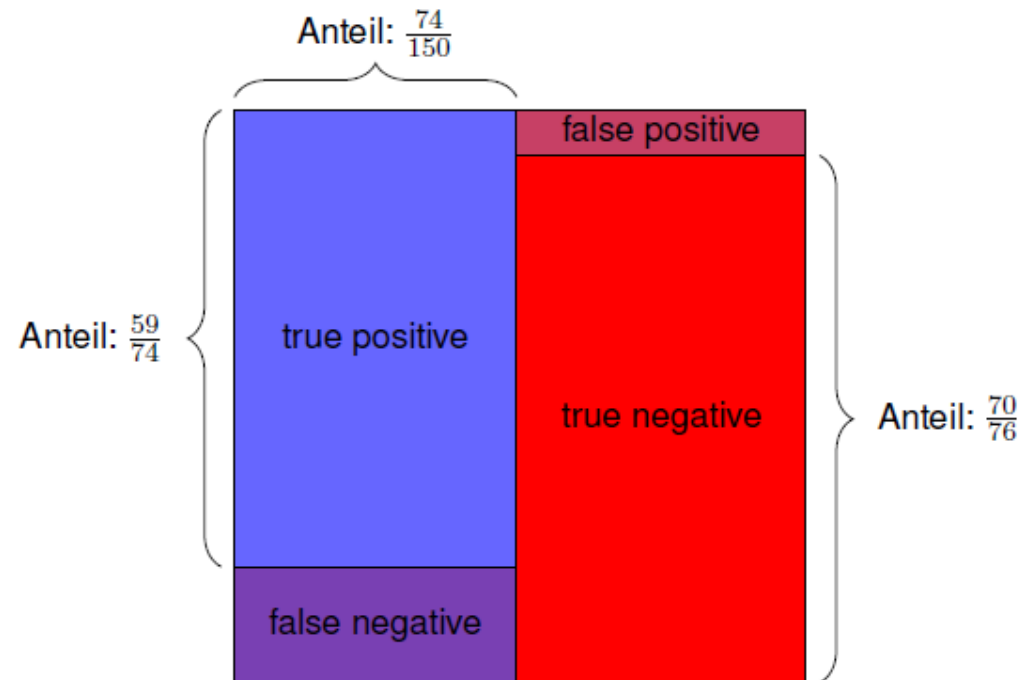
Konvention: Blau, also „gut“, ist Klasse 1, also die „positive“-Klasse.



# Die Confusion-Matrix und das Einheitsquadrat

Eine Visualisierung der Confusion-Matrix ist durch das Einheitsquadrat gegeben:

<i>absolut</i>	Kl. 1	Kl. 2	$\Sigma$
Kl. 1	59	6	65
Kl. 2	15	70	85
$\Sigma$	74	76	150



Die Größe der Felder spiegelt den prozentualen Anteil der jeweiligen Beispielgruppe an allen Beispielen wider.

# Accuracy

Das Gütemaß Accuracy entspricht einer Trefferquote des Modells, beantwortet also die Frage: „Wieviel Prozent der Beispiele in der Testmenge wurden durch das Modell richtig zugeordnet?“

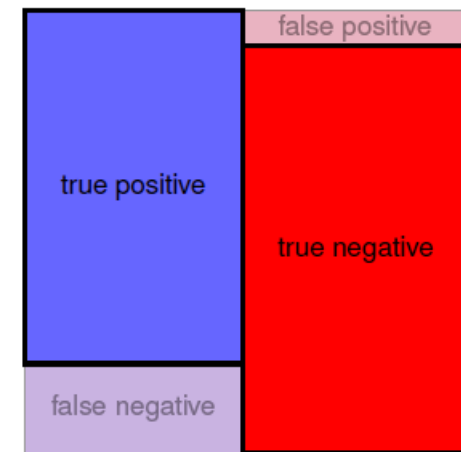
Accuracy

$$= \frac{\text{Richtig zugeordnete Beispiele}}{\text{Alle Beispiele}}$$

$$= \frac{tp + tn}{tp + fn + fp + tn}$$

$$= \frac{129}{150} = 0,86$$

<i>absolut</i>	Kl. 1	Kl. 2	$\Sigma$
Kl. 1	59	6	65
Kl. 2	15	70	85
$\Sigma$	74	76	150



# Accuracy im Vergleich

---

## Arbeitsauftrag:

Auf der nächsten Folie ist das Ergebnis der Testphase von zwei Klassifikationen dargestellt.

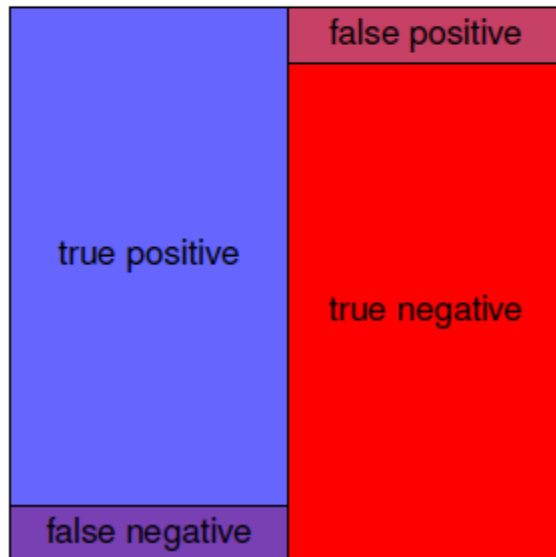
Diskutieren und notieren Sie die Antworten auf die folgenden Fragen. Nutzen Sie hierfür Mural (Link folgt in Veranstaltung).

- Wie groß ist die Accuracy der beiden Modelle?
- Für wie „gut“ beurteilen Sie die beiden Modelle und warum? Denken Sie dabei zurück an Ihre Überlegungen zur Modellqualität von Anfang.
- Welche Schlussfolgerungen lassen sich daraus für das Gütemaß Accuracy ziehen? Begründen Sie Ihre Antwort.

# Accuracy im Vergleich

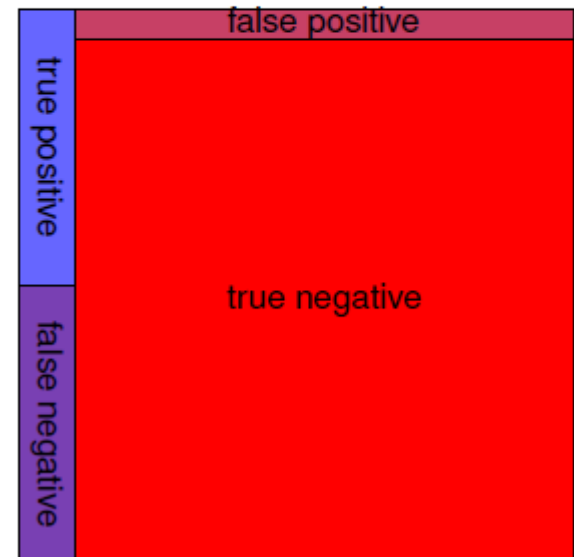
Modell A:

<i>absolut</i>	Kl. 1	Kl. 2	$\Sigma$
Kl. 1	90	10	100
Kl. 2	10	90	100
$\Sigma$	100	100	200



Modell B:

<i>absolut</i>	Kl. 1	Kl. 2	$\Sigma$
Kl. 1	10	10	20
Kl. 2	10	170	180
$\Sigma$	20	180	200



# Accuracy im Vergleich - I

---

Möglichkeiten:

- ❶ Klassifizierer A erkennt jedes Objekt aus der Testmenge mit einer Wahrscheinlichkeit von 90% richtig, unabhängig davon, aus welcher Klasse es stammt. Je nach Fragestellung kann das ein sehr guter Wert sein.
- ❷ Klassifizierer B erkennt nur Objekte der Testmenge, die aus Klasse 2 stammen sehr zuverlässig (>94%), erkennt Elemente der Klasse 1 aber gar nicht (Trefferwahrscheinlichkeit entspricht Ratewahrscheinlichkeit).

# Accuracy im Vergleich - II

---

Möglichkeiten:

- ❶ Klassifizierer sind nicht grundsätzlich bezüglich ihrer Accuracy vergleichbar. Die Accuracy muss immer in Zusammenhang zur Verteilung der Klassen innerhalb der Testmenge gesetzt werden. Außerdem enthält die Accuracy nicht grundsätzlich Informationen darüber wie gut eine bestimmte Klasse erkannt wird.

# Precision

Precision

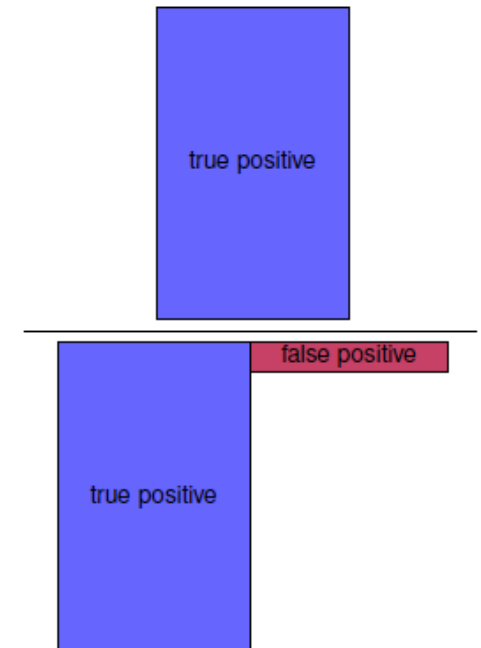
$$= \frac{\text{Richtig zugeordnete Beispiele von Kl. 1}}{\text{Alle Kl. 1 zugeordneten Beispiele}}$$

$$= \frac{tp}{tp + fp}$$

$$= \frac{59}{65} \approx 0,9077$$

Wünschenswert ist eine Precision von 1.

<i>absolut</i>	Kl. 1	Kl. 2	$\Sigma$
Kl. 1	59	6	65
Kl. 2	15	70	85
$\Sigma$	74	76	150





# Recall

Recall

$$= \frac{\text{Richtig zugeordnete Beispiele von Kl. 1}}{\text{Alle Beispiele von Kl. 1}}$$

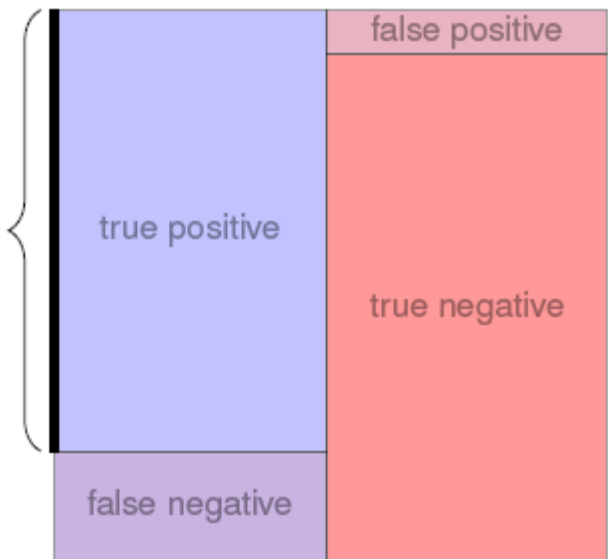
$$= \frac{tp}{tp + fn}$$

$$= \frac{59}{74} \approx 0,7973$$

Wünschenswert ist eine Recall von 1.

<i>absolut</i>	Kl. 1	Kl. 2	$\Sigma$
Kl. 1	59	6	65
Kl. 2	15	70	85
$\Sigma$	74	76	150

Anteil:  $\frac{59}{74}$



# Precision-Recall-Dilemma, F1-Score

- Hohe Precision: Vermeidung von „negative“-Beispielen, die „positive“ klassifiziert werden ( $fp$  wird minimiert).  
Nebeneffekt: Anzahl der „positive“-Beispiele, die „negative“ klassifiziert werden, steigt ( $fn$  wird größer).
- Hoher Recall: Vermeidung von „positive“-Beispielen, die der „negative“ Klasse zugeordnet werden ( $fn$  wird minimiert).  
Nebeneffekt: Anzahl der „negative“-Beispiele, die „positive“ klassifiziert werden, steigt ( $fp$  wird größer).
- Gibt die Fragestellung keine Präferenz, liefert der  $F_1$ -Score als Mittel zwischen Precision und Recall ein weiteres Gütemaß:

Meist: Je höher Precision, desto niedriger Recall und umgekehrt.

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

## 4.3 Varianz und Bias

---

# Modelleigenschaften

---

Neben den Gütemaßen gibt es noch eine weitere Modelleigenschaften, die es zu beachten gilt, und zwar **Varianz** und **Bias**.

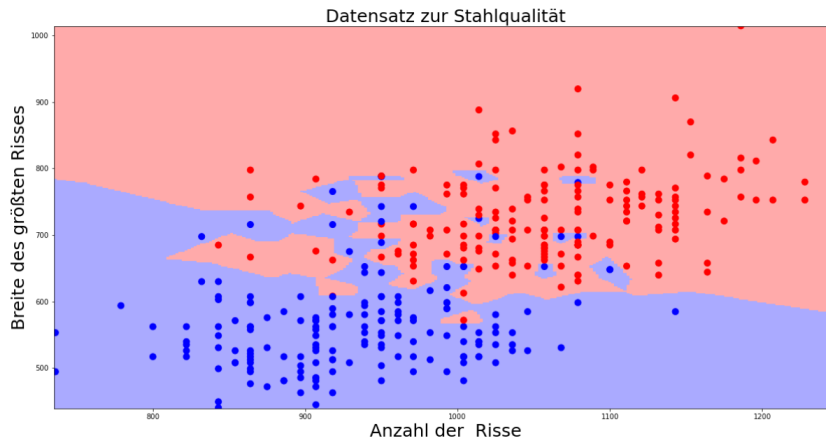
## **Hohe Varianz:**

Das Modell verändert sich stark, wenn eine andere Trainingsmenge (mit gleicher Verteilung) zum Trainieren genutzt wird.

## **Hohe Bias:**

Das Modell weicht im Mittel über alle möglichen Trainingsmengen (mit gleicher Verteilung) vom erwarteten Verhalten ab.

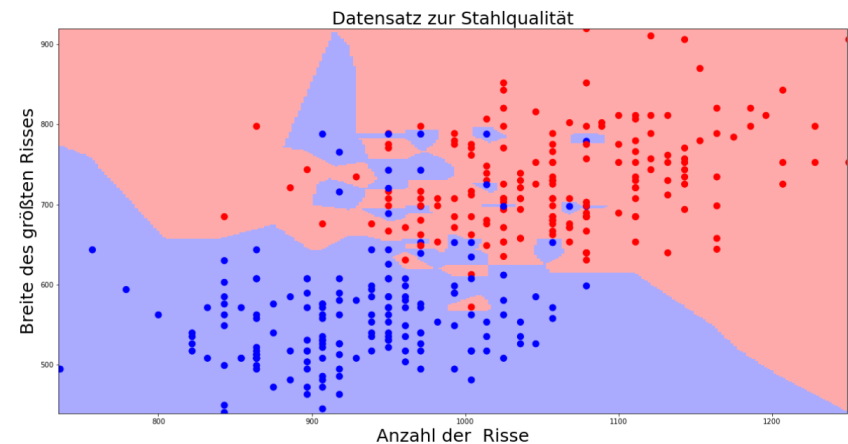
# Modell mit hoher Varianz



Modellparameter:  
k = 1, euklidische Metrik,  
Trennungparameter:  
randomstate=42

Modellparameter:  
k = 1, euklidische Metrik,  
Trennungparameter:  
randomstate=1

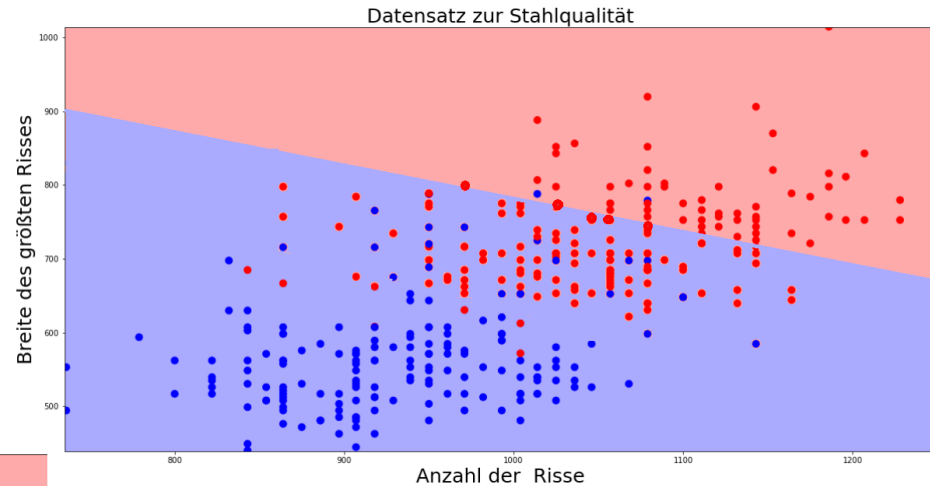
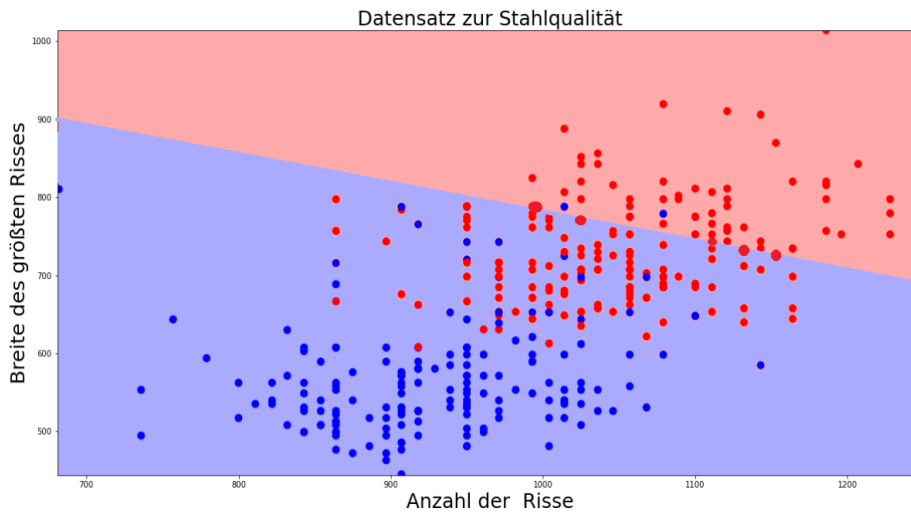
Starke Veränderung durch eine andere Trainingsmenge.



# Modell mit hohem Bias

Das Modell weicht im Mittel über alle möglichen Trainingsmengen (mit gleicher Verteilung) vom erwarteten Verhalten ab.

**Erwartetes Verhalten:**  
Rot auf rot und blau auf blau!



Hinweis: Ein solches Modell ist nicht mit dem kNN erzeugbar.

# Bias-Varianz-Dilemma

---

Das Ziel sollte stets sein, ein Modell mit geringer Varianz und geringem Bias zu konstruieren.

Leider stellt sich heraus, dass Bias und Varianz nicht unabhängig sind, sondern in den meisten Fällen ein Dilemma: Je niedriger die Varianz eines Modells, desto höher der Bias und umgekehrt.

→ Im Fall des kNN kommt dies nur vor, wenn  $k$  so hoch gewählt wird, dass die Fläche nur noch eine Farbe hat.

# Overfitting und Underfitting

---

**Overfitting (Überanpassung) und Underfitting (Unteranpassung)** bedeutet, dass ein Modell zu viel (**overfittet**) oder zu wenig (**underfittet**) auf die Daten angepasst ist.

**Overfitting führt zu Varianz**, denn je mehr trainiert wird, desto spezifischer wird das Modell auf die Trainingsdaten angepasst und eine andere Trainingsmenge führt so zu einem anderen Ergebnis.

**Underfitting führt zu Bias**, denn wenn das Modell noch nicht gut trainiert ist, führt es für jede Trainingsmenge nicht zu einem guten Ergebnis.



# 5 Projektabschluss

---

5.1 Antwort auf das Stahlbeispiel

5.2 Ausblick



Foto von Martinelle auf Pixabay

# Lernziele

---

1. Sie wenden Güteanalysen an und interpretieren die Ergebnisse in Bezug auf das Modell und auf die übergeordnete Fragestellung.
2. Sie optimieren Parameter der Verfahren in Hinblick auf die Gütemaße und die übergeordnete Fragestellung.
3. Sie nutzen ein in Python gefittetes Modell zur Klassifizierung eines neuen Datensatzes.

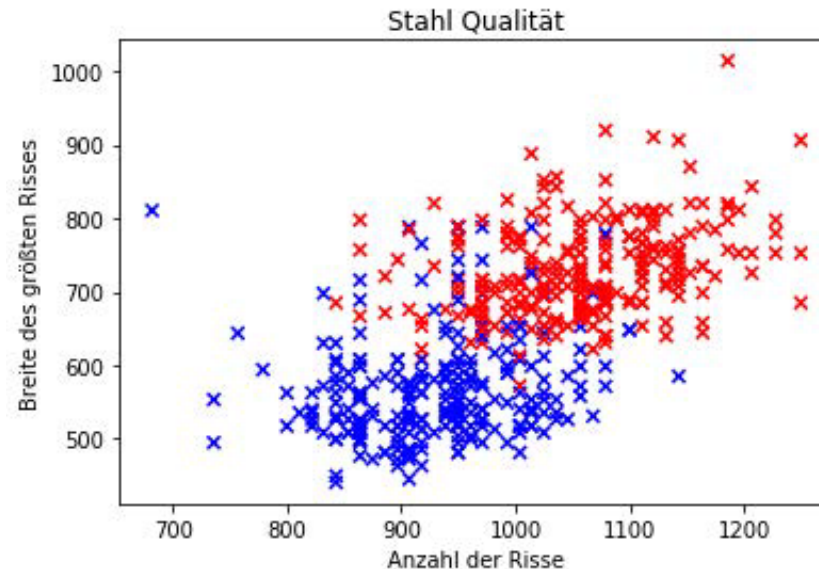
# 5.1 Antwort auf das Stahlprojekt

---

# Stahlprojekt: Modellerstellung

---

Es wird nun erneut das Ausgangsbeispiel zur Stahlqualität, sowie die Möglichkeit der Bibliothek scikit-learn für Maschinelles Lernen in Python betrachtet.



# Wiederholung:

## Modellerstellung überwachtes Lernen

---

- Mit dem überwachten Lernverfahren wird ein Modell (Zuordnung von Features zum Label) entwickelt.
- Trennung in Trainings- und Testdaten:
  - Mit Trainingsdaten wird das Modell entwickelt („gefittet“).
  - Mit Testdaten wird die Güte des Modells überprüft (Weicht das vom Modell vorhergesagte Label vom bereits bekannten Label in den Testdaten ab?).

# Modellvalidierung

## Stahlprojekt

---

Nächste Schritte:

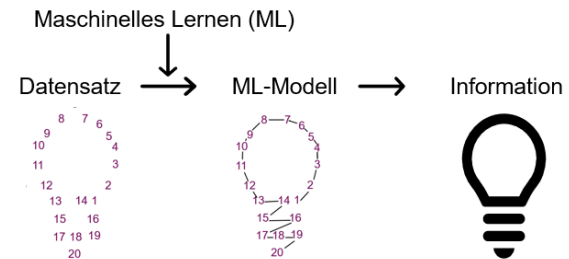
- Starten der Anaconda Distribution.
- Öffnen des Jupyter Notebooks „Modellvalidierung.kNN“.

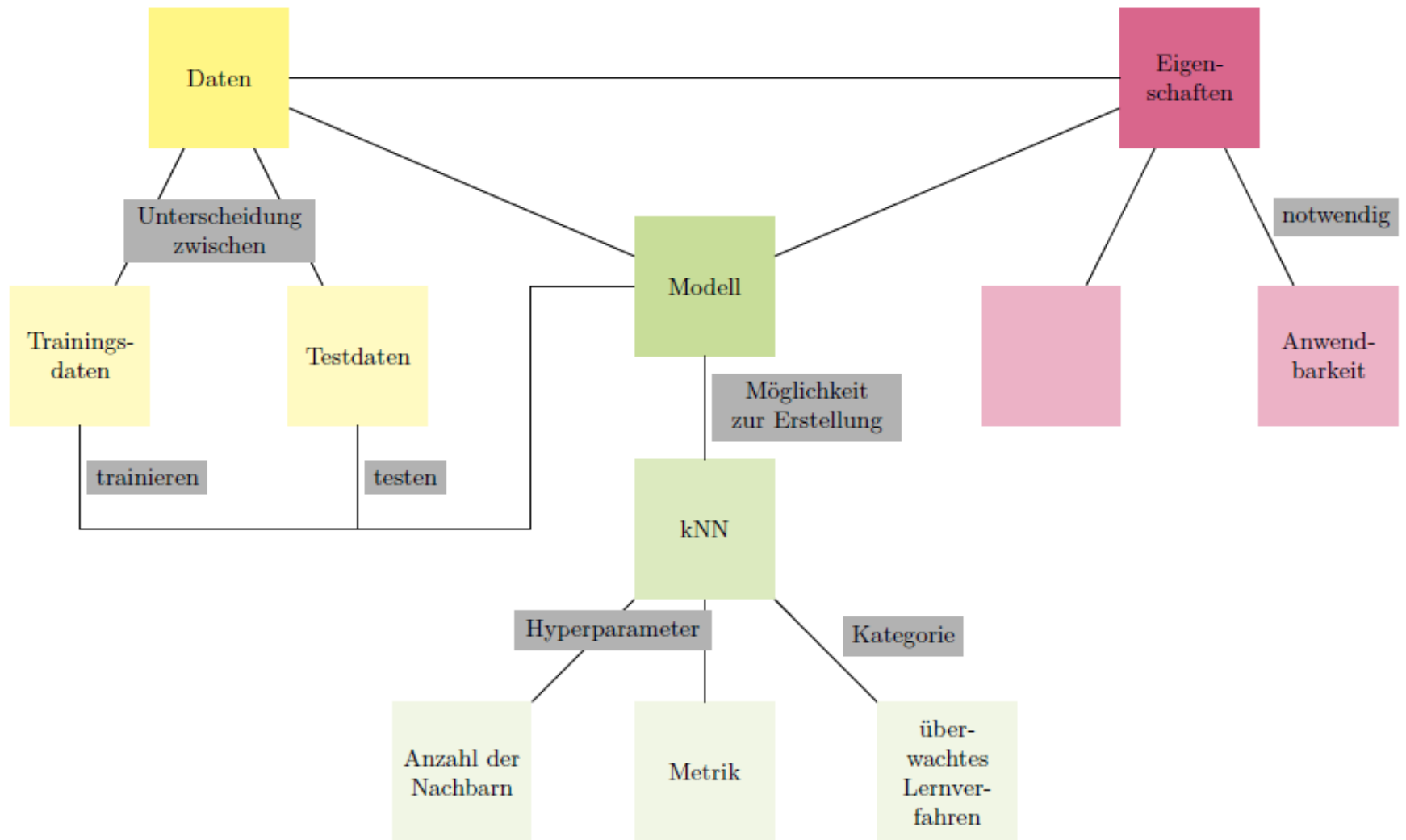


# Aufgabe zum Abschluss

Bearbeiten Sie mit Rückblick auf Die heutige Veranstaltung die drei folgenden Arbeitsaufträge:

1. Beschreiben Sie in 1-2 Sätzen, was ein ML-Modell im Kontext überwachtes Lernen ist.
2. Notieren Sie 3 Aspekte, an denen Sie die Qualität eines Modells beurteilen würden.
3. Beschriften Sie die nachstehende Grafik so, dass der Zusammenhang zwischen den einzelnen Begriffen durch die Grafik dargestellt wird. Erweitern Sie die Grafik um weitere in diesem Kontext relevante Begriffe und Verbindungen.







# 5.2 Ausblick

---

# Ausblick

---

- Die kNN-Klassifikation lässt sich mit mehr als zwei Klassen und mit mehr als 2 Features durchführen. Eine Durchführung mit mehr als 3 Features lässt sich nicht mehr gut visualisieren.
- Es gibt viele andere Verfahren des überwachten Maschinellen Lernens, mit denen Modelle gebildet werden können.

# Vielen Dank für Ihre Aufmerksamkeit!

---

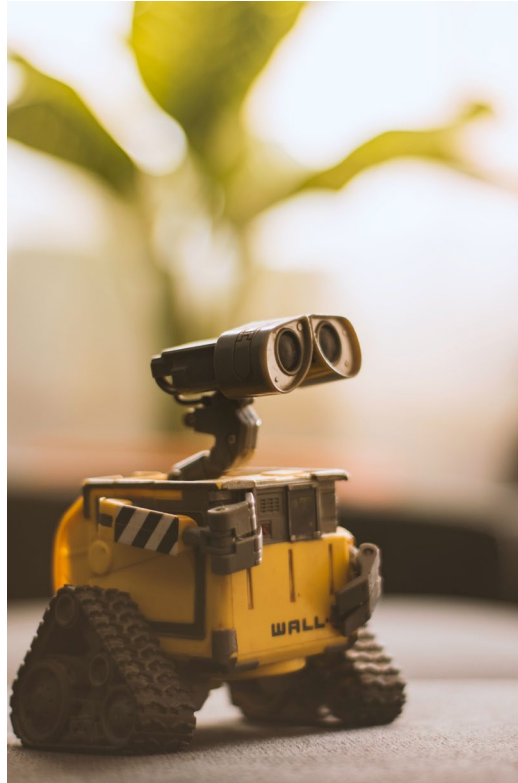


Foto von Lenin Estrada von Pexels