

Lage, Lage, Lage

Ein Machine Learning Ansatz

JOHANNES R. KOPKA

Zur Person

Johannes R. Kopka

2021 M.Eng. Geodäsie Hochschule Bochum

Masterarbeit zum Potenzial von Machine-Learning-Algorithmen bei der Anwendung auf Kaufpreissammlungen

2022 mit Prof. Dietmar Weigt: „Potenziale von Machine Learning Methoden bei der Auswertung der Kaufpreissammlungen“. In: fub – Flächenmanagement und Bodenordnung 3/2022 Juni, S. 120–128.

2023 Verm.Ass. in der Abt. Grundstückswertermittlung der Stadt Köln

M.Eng. Geoinformatik Hochschule Bochum

Masterarbeit zur Modellierung von Mietpreisen mittels Machine-Learning-Algorithmen



Das Problem: Was ist eine gute Lage?

Unterschiedliche (demographische) Gruppen präferieren verschiedene Lagen → **subjektives Merkmal**

In der Wertermittlung oft mittels intersubjektiver Betrachtungsweise

Gute Lage kann man oft nicht „sehen“ bzw. am Stadtbild festmachen

Beispiel Kölner Mietspiegel: „Die **sehr guten Wohnlagen** sind durch **aufgelockerte**, in der Regel **zweigeschossige** [...] Bebauung in **ruhiger** und verkehrsgünstiger **Grünlage** gekennzeichnet.“

→ wird man in hochpreisigen Innenstadtlagen kaum finden!

Als **praktisches** Merkmal im Rahmen der **Immobilienbewertung** ist der hier verfolgte Ansatz, die Lage im Zusammenhang mit dem **Wert (Preis)** zu betrachten

→ das Merkmal Lage soll (Miet-) **Preise erklären**

→ gute Lage = hoher Preis



*Straßenzug im hochpreisigen
„Belgischen Viertel“ in Köln
Quelle: Google Streetview*

Beispiel Köln: Wo ist eine gute Lage?

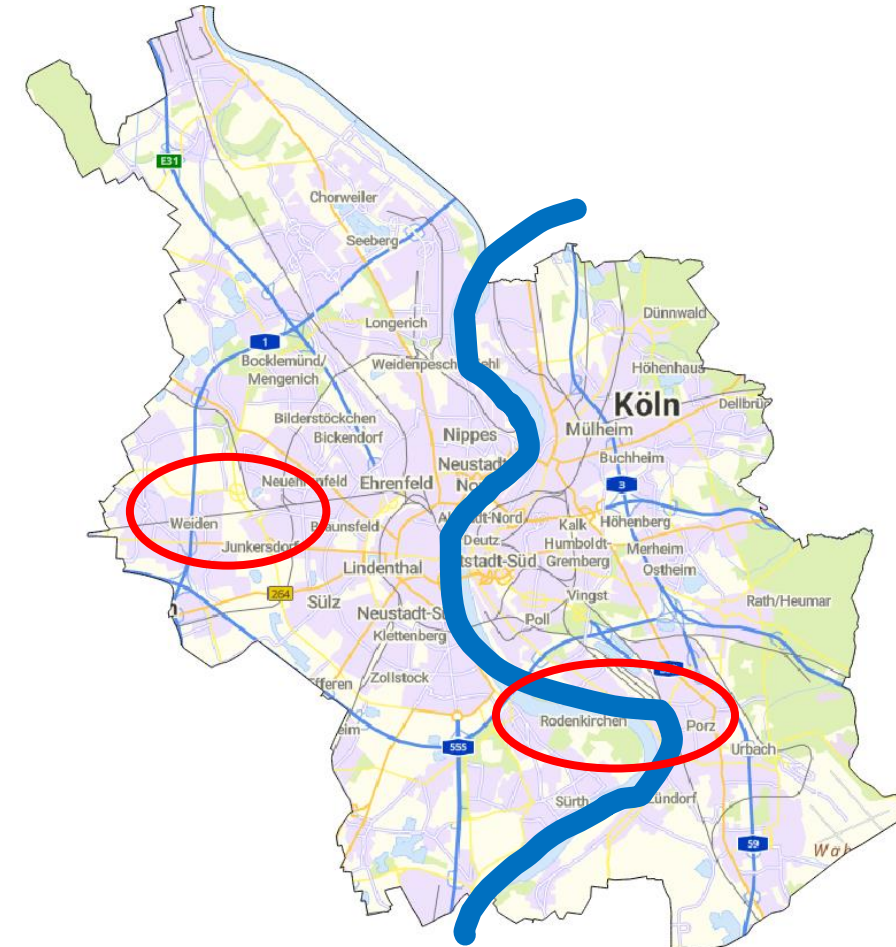
Größtes Stadtgebiet in Nordrhein-Westfalen

Umfangreiche Eingliederung benachbarter Gemeinden und Städte durch Köln-Gesetz von 1974

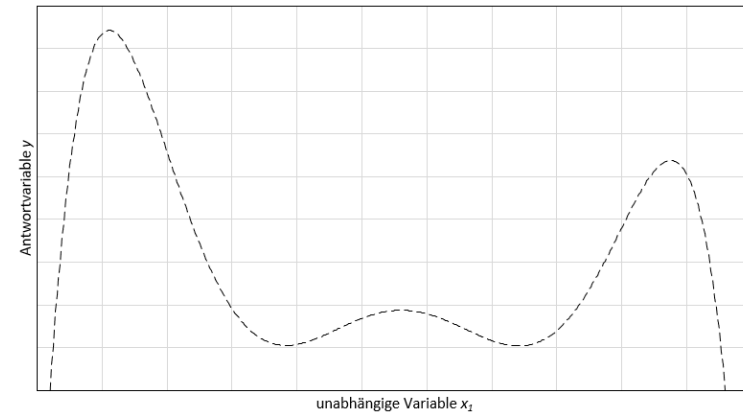
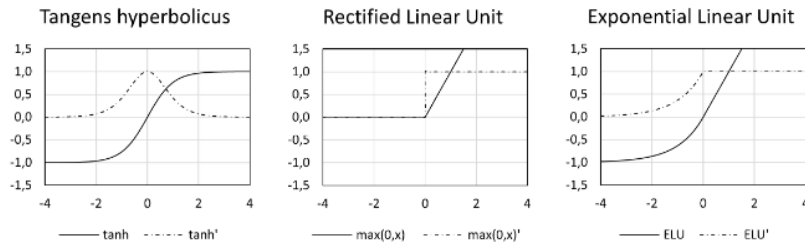
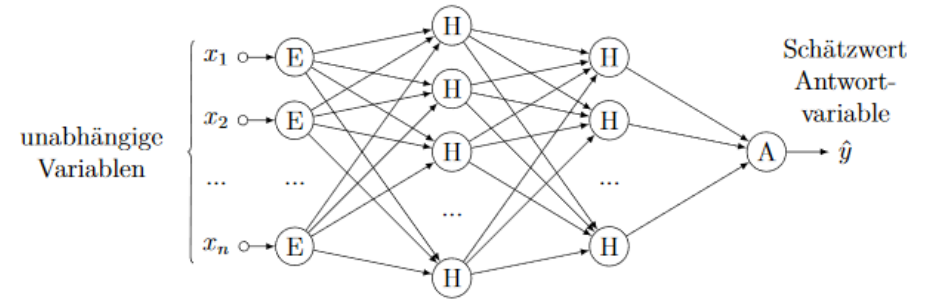
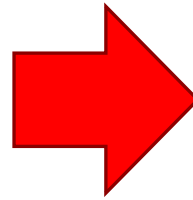
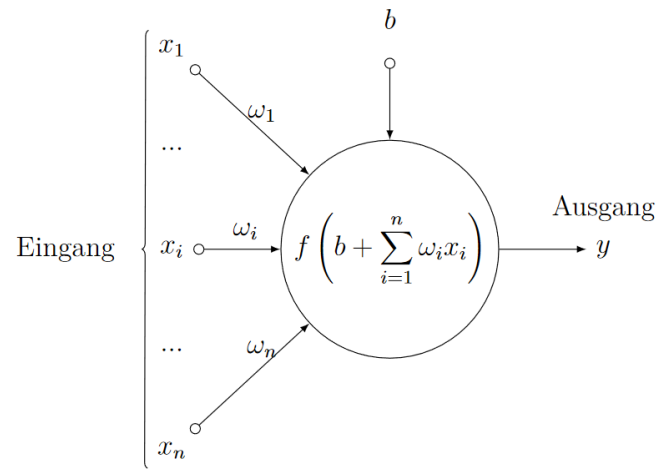
Als natürliche Barriere spaltet der Rhein die Stadt in eine linke und rechte Rheinseite

Attraktive Stadt mit teils sehr hohem Miet- und Preisniveau

→ äußerst heterogenes Lagebild auf dem Immobilienmarkt

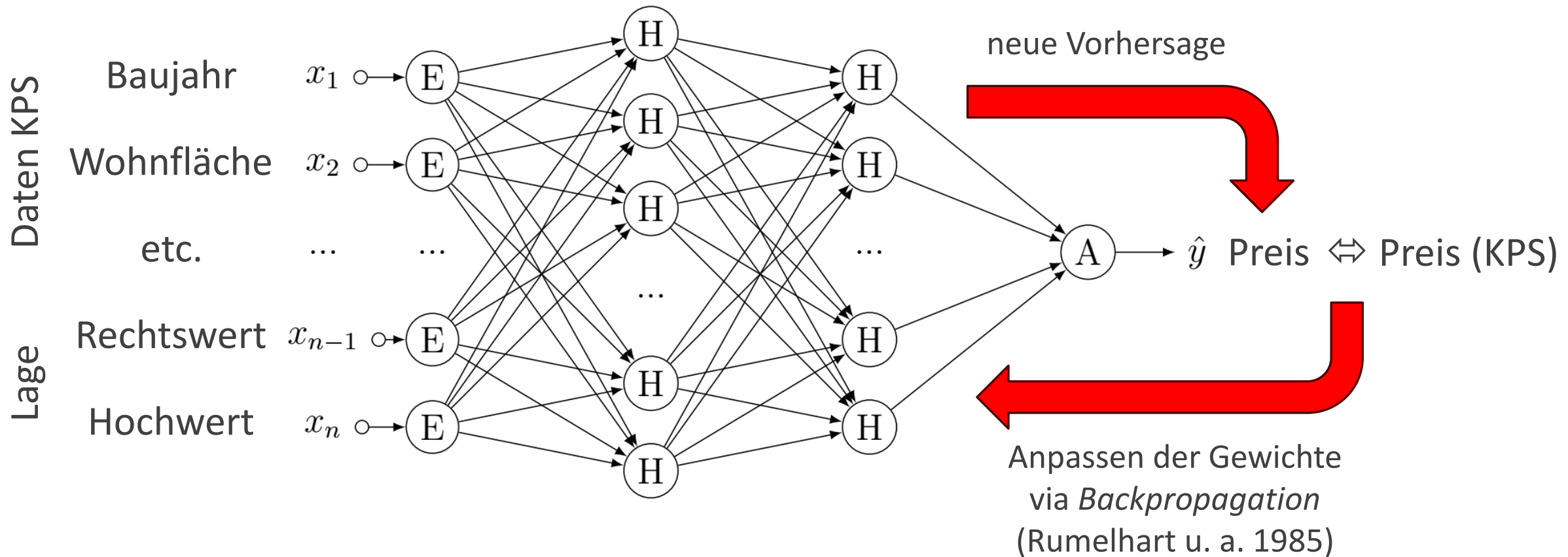


Vorteil neuronaler Netze

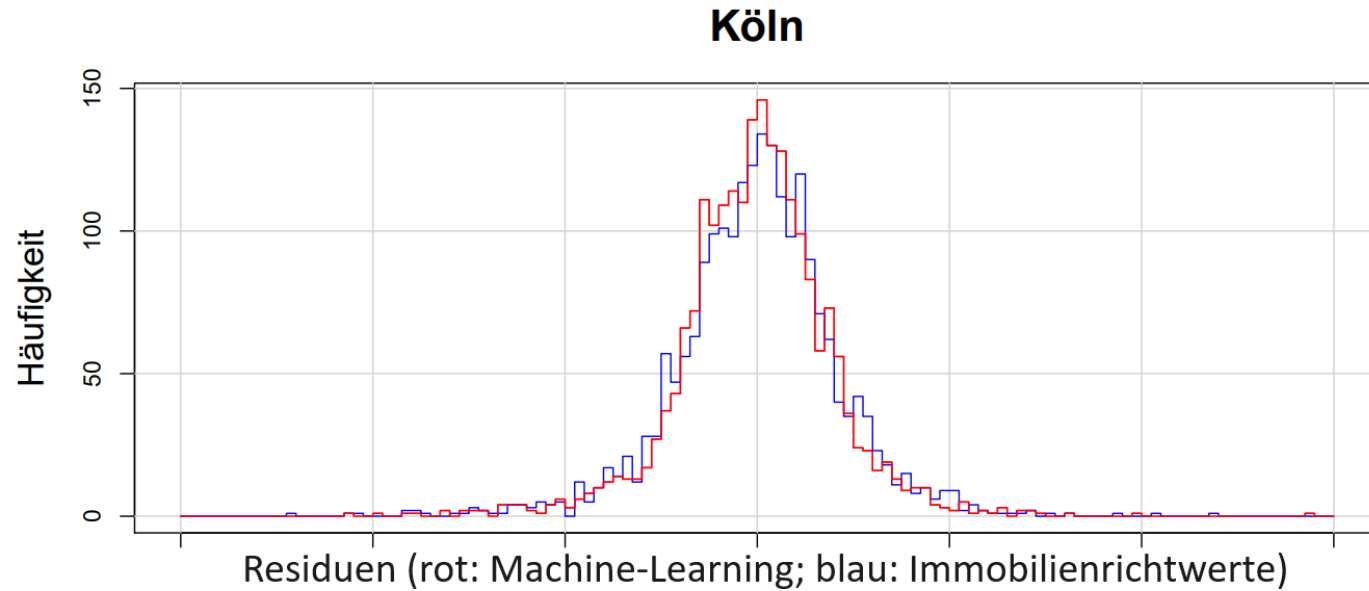


„Universal Approximation Theorem“ (s. Hornik 1991 und Leshno u. a. 1993)

Mustererkennung Beispiel Immobilien

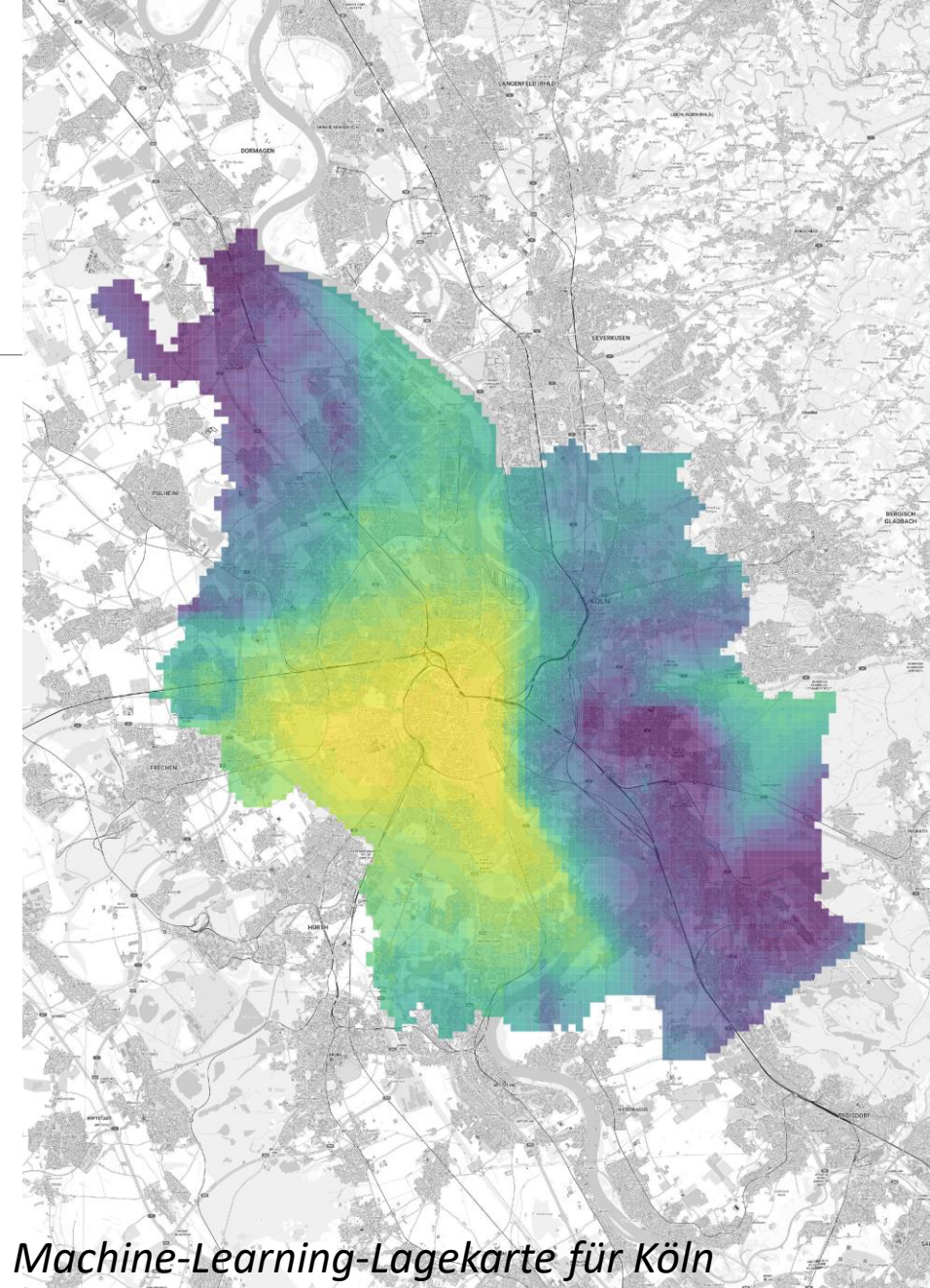


Ergebnisse Köln



Evaluierung u. a. durch Vergleich mit Immobilienrichtwerten
→ Machine Learning zeigt leicht aber signifikant kleinere Residuen (großes Rauschen auf dem Immobilienmarkt).

Komplexe 2D-Lagefunktion kann approximiert werden!



Machine-Learning-Lagekarte für Köln

Lagetransfer

Am Beispiel Köln konnte gezeigt werden, dass die Lage über das Koordinatenpaar dargestellt werden kann.

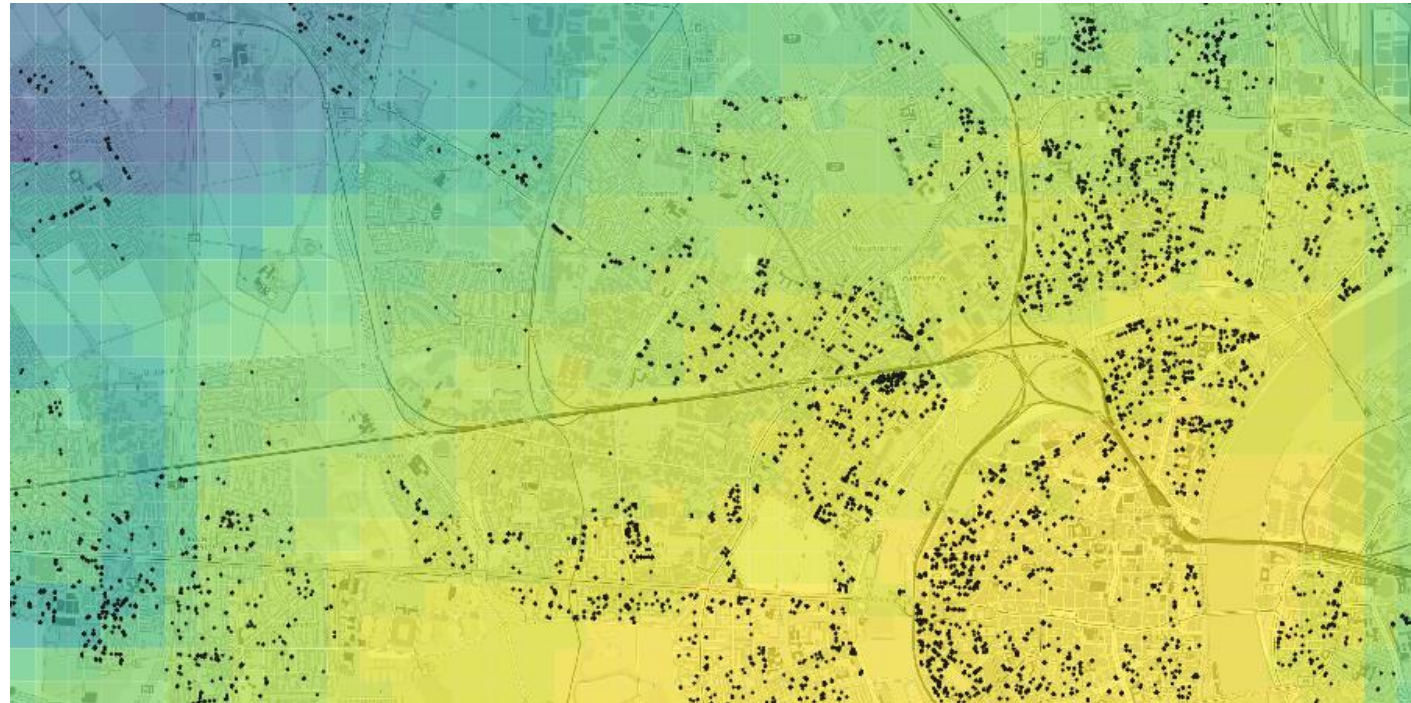
Koordinaten an sich haben aber keinen Einfluss auf die Lagequalität – sie beschreiben nur eine geographische Position

Zwischen und außerhalb der Trainingsdaten wird also inter- bzw. extrapoliert

Dies schränkt die Übertragbarkeit des Lagemodells ein

Idee: Koordinaten durch Geodaten substituieren, welche mit der Lage zumindest korrelieren

*Kölner Machine-Learning-Lagekarte
mit Darstellung der Trainingsdaten*



Koordinaten substituieren

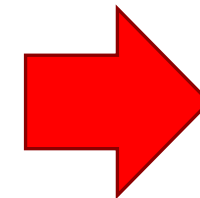
Zur Substitution der Koordinaten werden die Immobilien- mit **Geodaten** verschnitten

Dabei können z. B. abgeleitet werden:

- Entfernungen zu POIs wie Haltestellen, Schulen, etc.
- Sozioökonomische Daten wie Einwohnerdichte, Einkommen, etc.
- Immissionen wie Straßen- und Schienenlärm bei Tag und bei Nacht

Das Merkmal Lage kann hierdurch deutlich gestärkt werden

Merkmal	Anteil
Standard	27,4%
Baujahr	23,0%
Datum	12,7%
y	8,4%
Grundstücksgröße	7,3%
Wohnfläche	7,3%
x	6,7%
Anbauweise	2,9%
Stellplatz	2,4%
EFH oder ZFH	1,8%



Merkmal	Anteil
Standard	20,0%
Baujahr	15,6%
Datum	6,7%
Dist. Problemimmobilie	3,6%
Dist. Grundschule	3,3%
Wohnfläche	2,7%
Dist. Sportanlage	2,4%
Grundstücksgröße	2,3%
Dist. Gymnasium	2,1%
Dist. Spielplatz	2,1%
Dist. ÖPNV	2,0%
Dist. Trendsportart	1,9%
Dist. Sekundarschule	1,9%
Dist. Bürgerzentrum	1,8%
Y	1,8%

... weitere Merkmale

Ergebnisse aus Gelsenkirchen

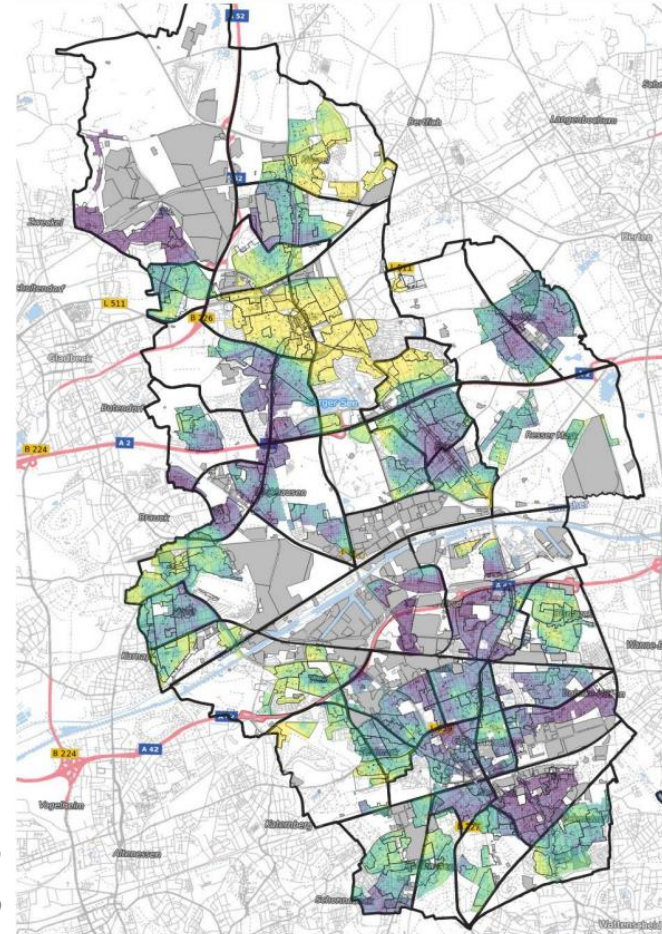
In Gelsenkirchen standen eine Vielzahl **hochwertiger Geodaten** zur Verfügung

Es konnte eine detaillierte Lagekarte (s. Abb.) erzeugt werden

Die Ergebnisse konnten von dem für Immobilienbewertung zuständigen Referat 62 plausibilisiert werden

In der Praxis zeigt sich ein hoher Nutzen der Lagekarte als Grundlage bei der Ableitung weiterer Produkte
→ „Mehrwert“

*Machine-Learning-Lagekarte Gelsenkirchen
aus Kopka und Weigt 2022*



Prinzipielle Unterschiede

HERKÖMMLICHE LAGEREPRÄSENTATIONEN

(Z. B. BODEN-/IMMOBILIENRICHTWEISEN)

Festlegungen a priori (z. B. Zonenbildung)

(inter)subjektive Einflüsse

i. d. R. keine Interaktionseffekte

nachvollziehbar

Wie löse ich das Problem?

MACHINE LEARNING

(Z. B. KÜNSTLICHES NEURONALES NETZ)

„parameterfrei“

möglichst autonom (objektiv)

universelle Approximation

„Black-Box“

Wie stelle ich das Problem?

→ Abwägung zwischen Transparenz und Performanz.

Machine Learning: Wie lässt sich das Problem formulieren, „wie lässt sich das Problem auf numerische Ein- und Ausgangswerte herunterbrechen“? (Kopka und Weigt 2022)

Literatur

Hornik, K. 1991: „Approximation capabilities of multilayer feedforward networks“. In: Neural networks 4.2, S. 251–257.

Kopka, J. und Weigt, D. 2022: „Potenziale von Machine Learning Methoden bei der Auswertung der Kaufpreissammlungen“. In: fub – Flächenmanagement und Bodenordnung 3/2022 Juni, S. 120–128.

Leshno, M. u. a. 1993: „Multilayer feedforward networks with a nonpolynomial activation function can approximate any function“. In: Neural networks 6.6, S. 861–8.

Rumelhart, D. E. u. a. 1985: Learning internal representations by error propagation. Techn. Ber. California University San Diego La Jolla Inst for Cognitive Science.

Ende

Vielen Dank für Ihre Aufmerksamkeit!