

MACHINE LEARNING VERBESSERT MIET- PREISVORHERSAGE UM BIS ZU 30 PROZENT

PRÄZISERE PREISBESTIMMUNG VON OBJEKTEN DANK MACHINE LEARNING

SITUATION AM MARKT

Mietpreise spielen beim Einschätzen der Rentabilität einer Immobilieninvestition eine große Rolle. Derzeit berechnen Investoren Mietpreise normalerweise aus öffentlichen Angeboten aus der Umgebung des Objektes, für das sie sich interessieren. Diese Art der Mietpreisermittlung bezeichnen wir als „Comparables“. Diese gängige Praxis führt aber häufig zu unpräzisen Ergebnissen, da Angebotsmieten stark nach Fläche, Baujahr und Zustand des Objektes variieren. Zudem stehen diese Faktoren in einem komplexen Zusammenhang mit der Lage des Objektes, wie zum Beispiel Anbindung, Grünflächen oder städtischer Infrastruktur. Hinzu kommt, dass in der näheren Umgebung keine vergleichbaren Angebote verfügbar sein können. Verlassen sich Investoren also nur auf aggregierte Durchschnitte bzw. Angebote aus der Umgebung, kann es bei der Ermittlung

von marktüblichen Preisen zu Fehlkalkulationen kommen. Eine Unterschätzung des Mietpreispotentials kann dazu führen, dass Mieteinnahmen zu gering angesetzt werden, was die erwartete Rendite schmälert und dass eigentlich solide Objekte für die Investition nicht in Betracht gezogen werden.

Maschinenlernalgorithmen helfen dabei, die marktübliche Ermittlung von Preisen im Vergleich zu Comparables zu präzisieren. Im Unterschied zu diesen sind Maschinenlernalgorithmen in der Lage, den Einfluss von Objekt- und Umgebungsvariablen sowie deren komplexe Interaktion auf die Mieten zu berücksichtigen.

METHODIK

Als Datengrundlage für die Studie der 21st Real Estate GmbH wird die Fläche Deutschlands in ein System aus über 55 Millionen Analysekecheln aufgeteilt aus denen 1,1 Millionen Ergebniskacheln generiert werden, die in Städten

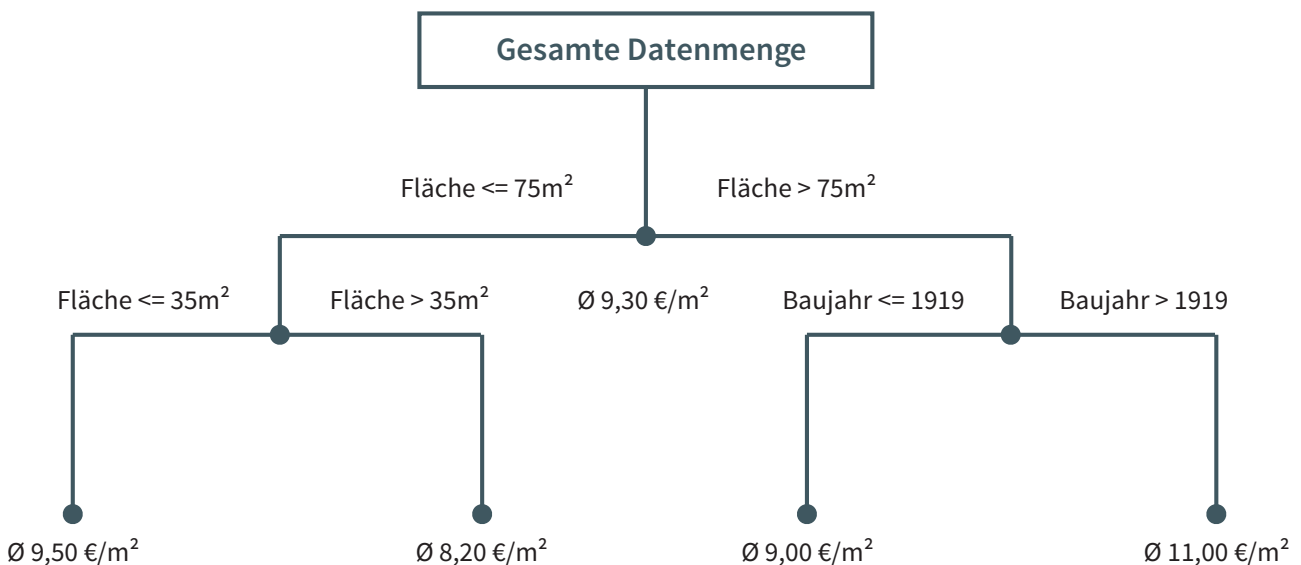
eine Größe von 200 × 200 Metern messen. Die Informationen aus diesen Kacheln werden mit unserer Datenbank aus über 40 Millionen Angeboten angereichert.

MACHINE LEARNING (RANDOM FOREST)

Der in der Studie verwendete Maschinenlernalgorithmus ist das sogenannte Random-Forest-Modell. Hierbei handelt es sich um einen Regressionsbaumansatz, bei dem Beobachtungen anhand von Einflussgrößen und Interaktionen zwischen diesen in homogene Untergruppen aufgeteilt werden. Diese Untergruppen werden immer feiner, bis entweder alle Beobachtungen denselben Wert haben oder nur fünf Beobachtungen in der Untergruppe verbleiben. Regressionsbäume können damit komplexe Zusammenhänge abbilden.

Die folgende Grafik veranschaulicht den Regressionsbaum an einem einfachen Beispiel. Für alle Daten werden in diesem Beispiel die Miete, das Baujahr und die Wohnfläche herangezogen. Im Anschluss werden Objekte auf Basis ihrer Eigenschaften in Gruppen mit ähnlichen Mieten sor-

tiert. Hierzu werden am ersten Knoten alle Objekte danach eingeteilt, ob sie weniger oder mehr als 75 Quadratmeter Fläche besitzen. Anschließend werden beide Untergruppen noch einmal nach dem Baujahr (rechts) und der Fläche (links) aufgeteilt. Die Durchschnittsmiete der hieraus resultierenden Gruppen ergibt die vorhergesagte Miete. Diese Bäume reichen sehr tief (mehr als 40 Tausend Knoten für Berlin und zwei bis drei Millionen Knoten für Gesamtdeutschland) und werden immer mit einem anderen Teil der Daten mehrfach wiederholt, wodurch sich die Aussagen besser generalisieren lassen.



COMPARABLES

Bei den Comparables werden die nächstgelegenen Angebote als Referenzbasis verwendet. Hierzu werden alle Angebote einer Stadt auf Kacheln (200 x 200 Meter) umgelegt und Durchschnittspreise pro Kachel berechnet. Ungesehene Objekte werden nun mit einem durch fußläufige Er-

reichbarkeit gewichteten Durchschnitt aller Kachelpreise in der nächsten Umgebung bewertet. Dabei werden Parameter wie die optimale Anzahl an Nachbarkacheln vorher optimiert.

PERFORMANCE

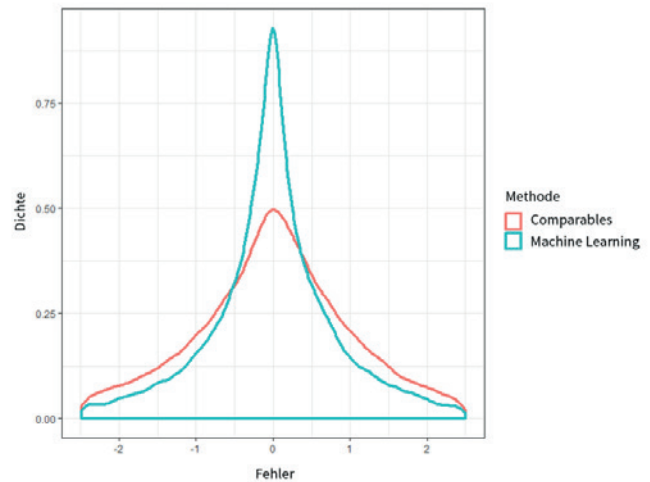
Um die Genauigkeit des Random-Forest-Modells und der Comparables zu bestimmen, wird der durchschnittliche absolute Schätzfehler (Mean Absolute Error) aller Vorhersagen der ungesehenen Objekte (in Euro pro Quadratmeter) berechnet. Um die Genauigkeit von Comparables und Random Forest zu veranschaulichen, wird für dasselbe Objekt der absolute Schätzfehler von beiden Methoden herangezogen und miteinander verglichen.

Das zugrundeliegende Verfahren ist eine sogenannte Kreuzvalidierung. Hierbei werden Modelle mit nur einem Teil der Daten trainiert um den nicht verwendeten („ungesehenen“) Teil der Daten zu schätzen. Die 21st Real Estate GmbH arbeitet mit einer 10-fachen Kreuzvalidierung, bei der immer 90 Prozent der Daten für das Training und 10 Prozent der Daten zur Evaluation verwendet werden. Die folgende Darstellung veranschaulicht das Verfahren:

Iteration 1	Test	Training	Training	Training	Training	Training	Training	Training	Training	Training
Iteration 2	Training	Test	Training	Training	Training	Training	Training	Training	Training	Training
Iteration 3	Training	Training	Test	Training	Training	Training	Training	Training	Training	Training
Iteration 4	Training	Training	Training	Test	Training	Training	Training	Training	Training	Training
Iteration 5	Training	Training	Training	Training	Test	Training	Training	Training	Training	Training
Iteration 6	Training	Training	Training	Training	Training	Test	Training	Training	Training	Training
Iteration 7	Training	Training	Training	Training	Training	Training	Test	Training	Training	Training
Iteration 8	Training	Training	Training	Training	Training	Training	Training	Test	Training	Training
Iteration 9	Training	Training	Training	Training	Training	Training	Training	Training	Test	Training
Iteration 10	Training	Training	Training	Training	Training	Training	Training	Training	Training	Test

ERMITTELTE SCHÄTZFEHLERVERTEILUNG

Der folgende Graph veranschaulicht die Verteilung des Schätzfehlers der Mieten für alle individuellen Angebote vom ersten Quartal 2017 bis einschließlich dem dritten Quartal 2017 in den ausgewählten Städten Berlin, München, Frankfurt am Main, Stuttgart, Hamburg, Düsseldorf, Köln, Leipzig, Nürnberg, Mannheim, Hannover, Essen und Bremen. Sowohl die Comparables als auch der Maschinenlernalgorithmus haben eine Verteilung die sich im Nullbereich zuspitzt, was einer fehlerlosen Schätzung entspricht. Somit weisen beide Verfahren keine systematische Verzerrung auf. Im Vergleich zu Comparables ist die Verteilung des Machine Learning-Fehlers jedoch wesentlich stärker um den Nullbereich konzentriert. Damit generiert der Maschinenlernalgorithmus Fehler mit einer konsistent niedrigeren Wahrscheinlichkeit.



ERGEBNIS

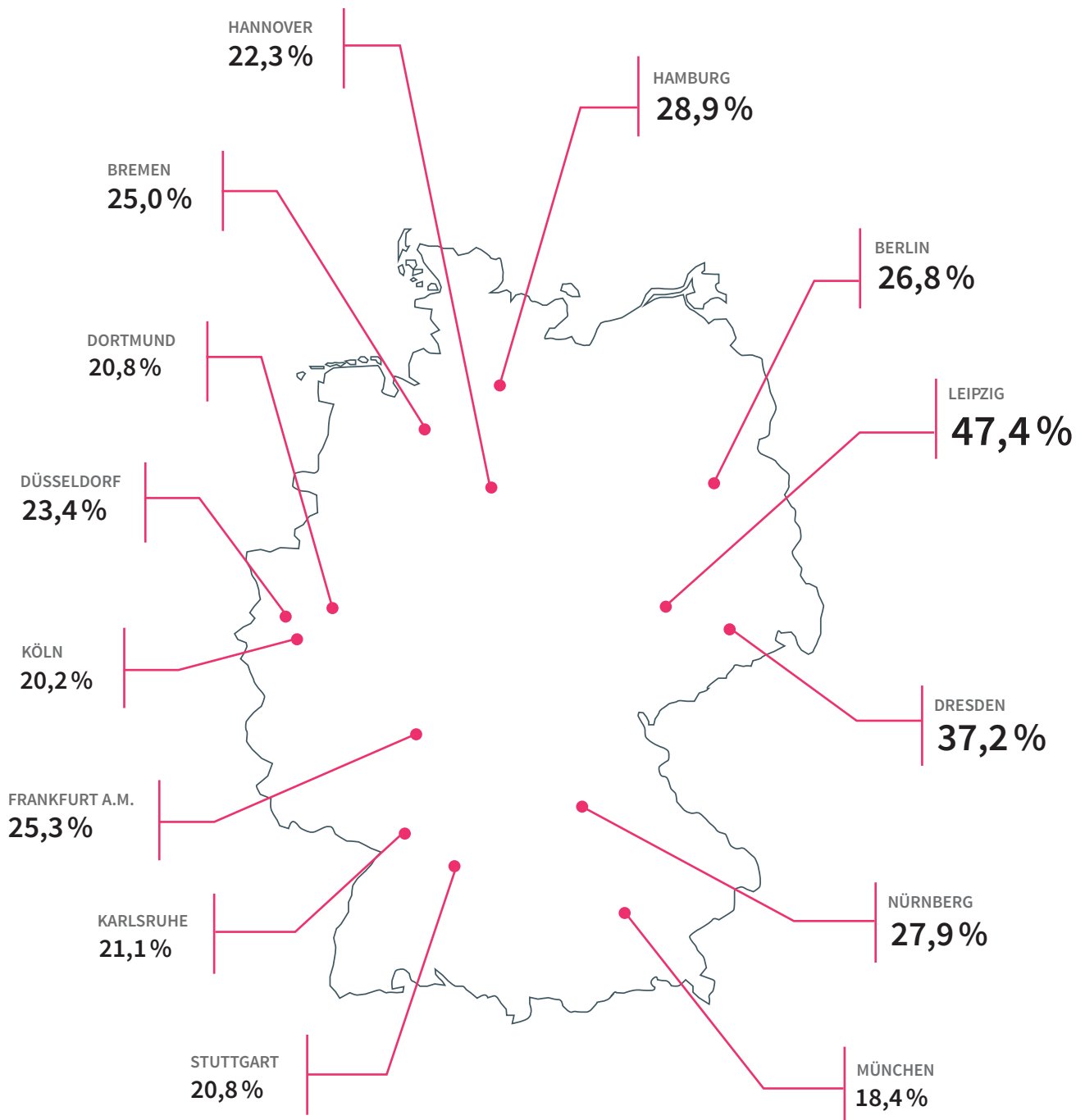
Die Studie zeigt, dass Mieten in A-Städten durchschnittlich um 25 Prozent genauer ermittelt werden, wenn Sie mithilfe von Maschinenlernalgorithmen anstatt mit lokalen Umgebungsdurchschnitten bestimmt werden. Der durchschnittliche absolute Schätzfehler liegt dabei bei 0,97 Euro pro Quadratmeter. So schätzt der Algorithmus Mietpreise in München um 18,4 Prozent genauer ein als ein herkömmliches Vergleichsverfahren. In Hamburg ist die Performance mit Machine Learning sogar um 28 Prozent genauer. In B-Städten performt der gewählte Maschinenlernalgorithmus sogar um 32 Prozent besser als Comparables mit einem durchschnittlichen absoluten Fehler von 0,48 Euro pro Quadratmeter. Hierbei ist der Maschinenlernalgorithmus für Leipzig mit 47,4 Prozent doppelt so genau als die Schätzung über Umgebungsmieten. Für Bochum, die Stadt mit der geringsten Verbesserung, ist das Machine Learning immer noch um 20 Prozent genauer. Mit einem durchschnittlichen absoluten Fehler von 0,55 Euro pro Quadratmeter verbessert der Maschinenlernalgorithmus die Ermittlung von Mieten in C-Städten um 28 Prozent.

Unsere Untersuchungen haben ergeben, dass Maschinenlernalgorithmen besonderes dann genauer als Comparables arbeiten, wenn die Charakteristika von Objekten (Baujahr, Fläche, Zustand) eine starke regionale Heterogenität aufweisen. Ferner steigt die Genauigkeit mit der Anzahl der zur Verfügung stehenden Objekte sowie der Streuung der Mieten.

Machine Learning ermöglicht eine präzisere Preisbestimmung von Objekten und somit eine verlässlichere Ermittlung von Renditezielen. Das Risiko von Leerstand oder einer zu geringen Rendite durch falsche Mietpreiserwartungen werden minimiert und Fehlinvestitionen vermieden.

Die Grafik von Deutschland zeigt, um wie viel Prozent die Mietpreisschätzung mittels Machine Learning präziser ist als Umgebungsrenten. Verdeutlicht am Beispiel von

Frankfurt am Main: Die Mietpreisschätzung mittels Machine Learning ist um 25,3 Prozent präziser als die Schätzung mit Comparables.



	REGION	VERBESSERUNG GGÜ. COMPARABLES	ABSOLUTER FEHLER MACHINE LEARNING IN €/m ²
A-STÄDTE	Hamburg	28,9%	0,9383
	Berlin	26,8%	0,8807
	Frankfurt am Main	25,3%	1,0358
	Düsseldorf	23,4%	0,9641
	Stuttgart	20,8%	1,2428
	Köln	20,2%	0,8922
	München	18,4%	1,3704
B-STÄDTE	Leipzig	47,4%	0,3262
	Dresden	37,2%	0,4229
	Nürnberg	27,9%	0,7452
	Wiesbaden	27,6%	0,7709
	Bremen	25,0%	0,6393
	Bonn	24,5%	0,6551
	Essen	24,1%	0,5011
	Mannheim	24,0%	0,7363
	Münster	23,9%	0,7983
	Duisburg	23,6%	0,3508
	Hannover	22,3%	0,8005
	Karlsruhe	21,1%	0,9682
	Dortmund	20,8%	0,4651
	Bochum	19,5%	0,5154

C-STÄDTE

REGION	VERBESSERUNG GGÜ. COMPARABLES	ABSOLUTER FEHLER MACHINE LEARNING IN €/m ²
Rostock	44,4%	0,3527
Erfurt	34,8%	0,4018
Potsdam	33,2%	0,4990
Saarbrücken	33,0%	0,4835
Magdeburg	31,2%	0,3566
Mönchengladbach	30,1%	0,4720
Kiel	27,5%	0,5990
Erlangen	27,3%	0,8515
Aachen	27,1%	0,6787
Darmstadt	26,0%	0,9505
Bielefeld	25,3%	0,6142
Wuppertal	24,7%	0,4931
Mainz	24,2%	0,7706
Lübeck	24,2%	0,5940
Osnabrück	24,1%	0,7723
Regensburg	19,0%	0,8714
Augsburg	19,0%	0,8153
Braunschweig	17,1%	0,7515
Heidelberg	16,0%	0,9855

ÜBER 21st REAL ESTATE GMBH

Die 2016 in Berlin gegründete 21st Real Estate GmbH ist ein Berliner Technologieunternehmen, das über Jahrzehnte IT- und Datenexpertise sowie über Transaktionserfahrung von mehreren Milliarden Euro verfügt. Mit diesem Know-

how entwickelt das Unternehmen innovative Produkte und IT-Services und transformiert Big Data zu relevanten Kennzahlen.

ANSPRECHPARTNER

Silvia Heinze
*Marketing & Communications
Manager*

Telefon: +49 (0)30 27 970 105
E-Mail: silvia.heinze@21re.de

Dr. Thorsten Martin
Senior Data Scientist

Telefon: +49 (0)30 94 868 314
E-Mail: thorsten.martin@21re.de