

# Maschinelles Lernen

## Modulararbeit – Aufgabenstellung

Prof. Dr. David Spieler, Prof. Dr. Stephanie Thiemichen  
Hochschule München

26. Juni 2024

## Aufgabenstellung – Generelles / Hinweise

Erstellen Sie einen Report, der Ihren Analyse-Workflow nachvollziehbar und reproduzierbar dokumentiert. Bearbeiten Sie die Aufgabenstellung mittels Python (konkret: Jupyter Notebook). Genauere Informationen zu den formalen Vorgaben (inklusive Umfang) finden Sie weiter unten unter "Formales". Bei den nachfolgenden Aufgaben, die zu bearbeiten sind, haben Sie alle Freiheiten hinsichtlich Struktur und Reihenfolge der Beschreibung, Wahl der Ansätze, Visualisierungen, etc.. Beachten Sie dazu die Hinweise zur Bewertung.

Gehen Sie beim Erstellen Ihres Reports davon aus, dass sich dieser unter anderem auch an Ihre Projektpartner:innen richtet, die nicht über tiefgreifendes Wissen im Bereich KI / Machine Learning verfügen, aber an der Vorbereitung der Datenerhebung beteiligt waren, sich in regelmäßigen Austausch mit Ihnen befinden und sehr an den Analyseergebnissen interessiert sind. Sie sollten gerade die wesentlichen Punkte also nachvollziehen können.

**Achtung:** Im Rahmen dieser Arbeit werden Sie an der ein oder anderen Stelle (vereinfachende) Annahmen treffen müssen. Dokumentieren Sie diese bitte entsprechend mit.

Versuchen Sie den Report nicht zu überladen, sondern konzentrieren Sie sich auf Kernaussagen, d.h. wesentliche Annahmen, Entscheidungen und Ergebnisse.

Code ist nach Bedarf und zur Gewährleistung von Nachvollziehbarkeit mit Kommentaren zu versehen.

## Bewertung

In der Modularbeit können Sie maximal 100 Punkte (ohne Bonus) erreichen. Die Punktevergabe erfolgt dabei anhand von drei Kategorien:

Kategorie	Beispiele	Max. Punkte
Umsetzung	Korrektheit, Nachvollziehbarkeit und Reproduzierbarkeit der verwendeten gewählten Algorithmen / Ansätze / Visualisierungen	40
Beschreibung	Korrektheit und Verständlichkeit des Reports	40
Stil	Sinnvolle Aufbereitung des Reports hinsichtlich Sprache, Grafiken, Code (inklusive Kommentaren), Struktur, ...	20

## Formales – Abgabemodalitäten

*Umfang:*

Die Länge Ihres Jupyter-Notebooks darf 400 Zeilen nicht überschreiten (unter Beibehaltung gängiger Style-Guide-Kriterien).

*Abgabe-Datei:*

Erstellen Sie ein ZIP-Archiv `ML_MA_Nachname_Vorname_Matrikelnummer.zip` mit folgendem Inhalt:

- i. `ML_MA_Nachname_Vorname_Matrikelnummer.ipynb`: Ihre Bearbeitung der Aufgaben als Jupyter-Notebook.

- Der File muss im 1. Code-Chunk sämtliche verwendete Pakete laden.
  - Der File muss am Ende im letzten Code-Chunk `sinfo()` (<https://pypi.org/project/sinfo/>) aufrufen und ausgeben.
  - Der File muss auf einem anderem System als dem Ihren lauffähig sein (lediglich die Pfadangabe zum Datenfile soll ggf. angepasst und die benötigten Pakete installiert werden müssen).<sup>1</sup>
- ii. `ML_MA_Nachname_Vorname_Matrikelnummer.html`: Die aus Ihrem Jupyter-Notebook kompilierte .html-Datei (für Hilfe siehe <https://ipython.org/ipython-doc/3/notebook/nbconvert.html>).
- iii. `ML_MA_Nachname_Vorname_Matrikelnummer_requirements.txt`: Spezifizieren Sie hier bitte die von Ihnen verwendeten Python-Packages inklusive deren Versionen, sowie die verwendete Python Version selbst.

### Abgabefrist

Reichen Sie das ZIP-Archiv via Moodle bis zum **17.07.2024 – Mitternacht** ein.

---

<sup>1</sup>Paketquellen: Beschränken Sie sich bei den zusätzlichen Paketen bitte auf Pakete, die via conda oder pip / PyPi verfügbar sind.

## Aufgabenstellung – Daten

Gegeben ist ein Teildatensatz der Daten des Münchener Mietspiegels aus dem Jahr 2023. Dieser enthält Angaben zu 3072 Mietwohnungen.

Die Datei `miete23.csv` finden Sie im Moodle der Veranstaltung neben dieser Aufgabenstellung.

**Bitte beachten:** Die Mietspiegeldaten 2023 sind nur im Rahmen dieser Lehrveranstaltung freigegeben und dürfen daher nicht weitergegeben werden!

### Beschreibung der Merkmale

Variablenname	Frage / Beschreibung
<code>nmqm</code>	Nettomiete pro $m^2$ (in €) der Wohnung
<code>wfl.gekappt</code>	Wohnfläche (in $m^2$ ); Kappung: Wohnfläche von Wohnungen unter 20 $m^2$ wurde auf 20 $m^2$ und über 160 $m^2$ wurde auf 160 $m^2$ gesetzt
<code>bj</code>	Baujahr des Gebäudes
<code>WL</code>	Wohnlage
<code>Zentral</code>	Liegt die Wohnung zentral in München? (1 = ja, 0 = nein)
<code>Gebtyp.23</code>	Gebäudetyp
<code>Modern.Boden</code>	Wurde der in allen Wohnräumen der Fußbodenbelag 2013 oder später modernisiert oder instandgesetzt? (1 = ja, 0 = nein)
<code>Boden.Fak</code>	Fußbodenbelag
<code>kue.score.23</code>	Anzahl zusätzlicher bereits vorhandener Küchenelektrogeräte (von 0 = keines bis max. 3), berücksichtigt werden können: Ceran- oder Induktionskochfeld, Kühlschrank oder Kühl-/Gefrierkombination, Geschirrspülmaschine
<code>bad.score23</code>	Score der angibt, wie viele der folgenden Merkmale vorliegen (von 0 = keines bis 4 = alle): Besondere Zusatzausstattung im Bad vorhanden, Handtuchheizkörper vorhanden, Fläche des (größten) Bades beträgt mindestens 6 $m^2$ und/oder zweites ausgestattetes Bad vorhanden, modernisiertes Bad
<code>Fußboden.Heizung</code>	Wohnung verfügt über Fußbodenheizung (1 = ja, 0 = nein)

### Hinweis:

Weitere Informationen zu den Variablen und zur Datenerhebung finden Sie unter <https://2023.mietspiegel-muenchen.de/dokumentation.html>.

### Aufgabe 1 (Überblick)

Laden Sie die Daten in ein Jupyter-Notebook und transformieren Sie die Daten in für Sie geeignete Formate (inklusive eventuell nötiger Vorverarbeitungsschritte für die spätere Modellierung). Verschaffen Sie sich einen Überblick über den Datensatz mit Hilfe von aussagekräftigen Visualisierungen (Diagramme, Tabellen, ...).

Gibt es Dinge, die Ihnen auffallen? Dokumentieren Sie Ihre Erkenntnisse.

### Aufgabe 2 (Modellierung)

Zielgröße einer Modellierung ist die Nettomiete pro  $m^2$ . Diese soll in Abhängigkeit der restlichen verfügbaren Variablen im Datensatz modelliert werden.

Wählen Sie zwei verschiedene, in der Vorlesung behandelte Modelltypen. Warum sind diese aus Ihrer Sicht für die Fragestellung geeignet? Beschreiben Sie kurz Vor- und Nachteile der einzelnen Methoden. Trainieren Sie die entsprechenden Modelle. Vergleichen Sie die Ergebnisse mittels geeigneter Metriken. Für welches Modell entscheiden Sie sich und warum? Ändert sich Ihre Entscheidung, wenn Interpretierbarkeit des Modells ein wesentliches Kriterium sein soll?

### Aufgabe 3 (Ergebnisinterpretation und Vorhersage)

Interpretieren Sie die Ergebnisse Ihrer Modellierung. Welche Variablen spielen eine Rolle?

Sagen Sie mittels des Modells die Nettomiete pro  $m^2$  für Ihre Wohnung voraus (wenn Sie in einem Wohnheim o.Ä. oder nicht in München wohnen, nehmen Sie einfach fiktive Werte an).