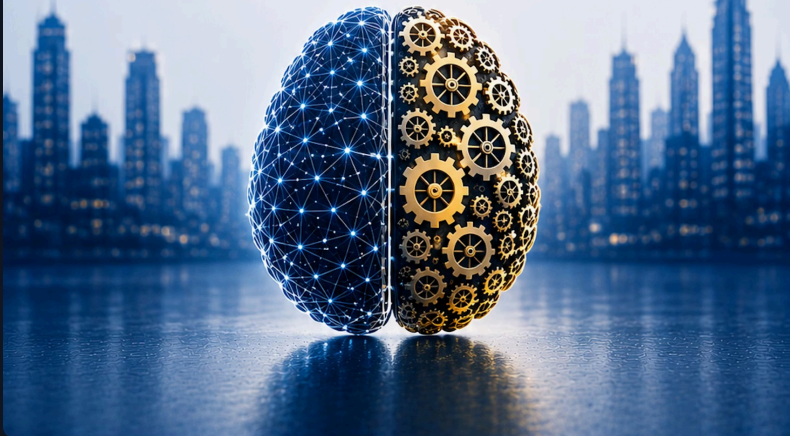


— MATHIAS ELLMANN —

VON MODELLEN ZU ENTSCHEIDUNGEN IN MACHINE LEARNING MIT PYTHON

Vom Modell zur
verantwortbaren Entscheidung



Von Modellen zu Entscheidungen in Machine Learning mit Python

Vom Modell zur verantwortbaren Entscheidung

Mathias Ellmann

ISBN: 978-3-6952-6182-6

Warum Machine Learning heute so wichtig ist

Machine Learning hilft, aus wachsender Datenmenge und Komplexität bessere Entscheidungen abzuleiten.

Datenmengen wachsen

Mehr Daten als je zuvor — intern, extern, strukturiert und unstrukturiert.

Komplexität steigt

Systeme und Märkte werden vielschichtiger; einfache Kausalitäten nehmen ab.

Entscheidungen werden schwieriger

Mehr Unsicherheit, mehr Tempo, höhere Konsequenzen.

Machine Learning als Antwort

Ein methodisches Werkzeug, um Orientierung in komplexen Situationen zu gewinnen.

Konsequenz: Ohne datenbasierte Unterstützung werden Priorisierung, Prognose und Risikoabwägung deutlich schwerer.

Das große Missverständnis

Machine Learning ist kein Modellwettbewerb — entscheidend sind bessere Entscheidungen.

Was viele fokussieren

- Algorithmen und neuronale Netze
- Modellkomplexität
- Accuracy und Benchmarks
- „Künstliche Intelligenz“ als Selbstzweck

Das eigentliche Ziel

Fundiert, nachvollziehbar und verantwortbar entscheiden.

Die Schlüsselfrage ist nicht: *Wie gut ist das Modell?* Sondern: *Was folgt daraus — und wer trägt die Konsequenzen?*

- Die Verschiebung vom Modell zur Entscheidung ist der Kern dieses Buches.

Die zentrale These des Buches

Der Weg von Daten zur Verantwortung ist keine technische Pipeline — er ist ein **Denkprozess**.



Entscheidend ist nicht das Modell, sondern die Konsequenz: bessere, nachvollziehbare Entscheidungen.

Warum Modelle benötigt werden

Modelle verdichten Komplexität und machen aus Daten handlungsfähige Orientierung.



Kognitive Grenzen

Menschen können nicht alle Daten direkt verarbeiten.



Muster erkennen

Modelle machen Strukturen in komplexen Daten sichtbar.



Prognosen ermöglichen

Sie stützen Aussagen über zukünftige Entwicklungen.



Entscheidungen absichern

Modelle unterstützen Urteile — sie ersetzen sie nicht.

Beispiel: Ohne Modell bleibt eine Analyse fragmentiert; mit Modell wird sie anschlussfähig für Entscheidungen.

Modelle sind nicht die Wirklichkeit

„Die Karte ist nicht das Gebiet.“ — Alfred Korzybski

Thesis: Modelle sind nützliche Vereinfachungen — nicht die Realität selbst.

Landkarte als Analogie

Eine Karte wählt aus, betont und lässt weg. Modelle tun dasselbe.

- lernen aus Beobachtungen, nicht aus der Wirklichkeit
- bilden nur Ausschnitte ab
- beruhen auf menschlicher Auswahl

Gefahr der Modellgläubigkeit

Wer Modelle mit Wahrheit verwechselt, trifft Entscheidungen mit falschem Sicherheitsgefühl.

- Modellaussagen sind kontextabhängig
- Übertragung auf neue Kontexte ist riskant
- Modelle sind Werkzeuge, keine Wahrheiten

Konsequenz: Ein gutes Modell hilft bei Orientierung — die Grenze zwischen Aussage und Wirklichkeit muss klar bleiben.

Jedes Modell vereinfacht

Vereinfachung ist notwendig. Entscheidend ist, **wie angemessen** sie ist.

Annahmen setzen Grenzen

Jedes Modell trifft Annahmen — und blendet dadurch manches aus.

Komplexität ist kein Qualitätsmerkmal

Am besten ist nicht das komplexeste, sondern das passendste Modell.

Ockhams Rasiermesser

So einfach wie möglich, so komplex wie nötig.

Beispiel: Ein überkomplexes Modell kann Daten nur auswendig lernen — und in neuen Fällen schlechter abschneiden.

Machine Learning als Mustererkennung

Kernthese: Machine Learning erkennt Muster in Daten und liefert Wahrscheinlichkeiten — keine Gewissheiten.

Was Machine Learning tut

- lernt **statistische Muster** aus Trainingsdaten
- verallgemeinert auf unbekannte Fälle
- arbeitet mit **Wahrscheinlichkeiten**

Konsequenz: Verantwortlicher Einsatz heißt, Ergebnisse als Unterstützung zu nutzen — nicht als Ersatz für Urteil und Expertise.

Was Machine Learning nicht tut

- es **versteht** keine Konzepte
- es ersetzt keine Fachkompetenz
- es liefert keine Gewissheiten

Warum Machine Learning keine Kausalität entdeckt

Korrelation ≠ Ursache

→ Muster ≠ Ursachen

Modelle lernen statistische Zusammenhänge — nicht Ursache-Wirkung.

→ Feature Importance ≠ Kausalität

Wichtig für Prognosen heißt nicht ursächlich.

→ Scheinkorrelationen

Zusammenhänge können zufällig sein, durch Drittvariablen entstehen oder Messfehler spiegeln.

→ Vorhersagen vs. Interventionen

Für Maßnahmen braucht es Kausalwissen — nicht nur Musterwissen.

Beispiel: Ein Modell erkennt, dass Eisverkauf und Badeunfälle zusammenhängen. Die Ursache ist nicht Eis, sondern heißes Wetter.

Vorhersagen sind keine Entscheidungen

ZENTRALE ERKENNTNIS DES BUCHES

Eine **Vorhersage ist kein Handlungsentscheid.** Sie liefert nur einen datenbasierten Hinweis.

Vorhersage

- Wahrscheinlichkeit auf Basis gelernter Muster
- Erwarteter Zustand unter bekannten Bedingungen
- Hinweis — kein Handlungsbefehl

Entscheidung

- **Ziele:** Was soll erreicht werden?
- **Werte:** Was zählt wirklich?
- **Kontext:** Für wen, unter welchen Bedingungen?
- **Urteilstkraft:** Die Integration aller Dimensionen

Konsequenz: Von der Vorhersage zur Handlung braucht es menschliche Urteilstkraft.

Wahrscheinlichkeiten sind keine Handlungsanweisungen

Ein Modell liefert Wahrscheinlichkeiten — keine Befehle.

Wahrscheinlichkeiten sind keine Garantien

80 % heißt auch: 20 % Fehlerrisiko — mit realen Folgen.

Verzerrte Wahrnehmung

Kleine Wahrscheinlichkeiten werden oft überschätzt, große unterschätzt.

Schwellenwerte sind Entscheidungen

Ab wann gehandelt wird, ist eine Wertentscheidung — keine technische Tatsache.

Mathematische Objektivität täuscht

Zahlen verdecken oft die menschlichen Wertentscheidungen dahinter.

Konsequenz: Von der Prognose zur Handlung braucht es explizite Urteilskraft.

Die Rolle menschlicher Urteilskraft

Modell + Urteilskraft = Entscheidung

Was Urteilskraft leistet

Sie verbindet Modellaussage und Entscheidung: Kontext, Erfahrung, Werte und situative Besonderheiten.

- Modelle unterstützen — sie ersetzen nicht
- Sie benennt, was nicht gemessen wird
- Sie trägt die Konsequenzen

Human-in-the-Loop

Der Mensch bleibt Teil des Systems. Entscheidungen brauchen Verantwortungsträger.

- Mensch als Kontrollinstanz
- Explizite Verantwortungszuweisung
- Verantwortungsvolle KI beginnt hier

Warum Modelle keine Lösungen sind

Lösungen verändern die Wirklichkeit — Modelle beschreiben sie nur.

Kontext entscheidet

Ein Modell kann in einem anderen Kontext scheitern — auch bei guten Kennzahlen.

Werkzeugfixierung

Das Modell wird zum Selbstzweck, statt das eigentliche Problem zu lösen.

Kennzahlen ≠ Ziele

Ein Modell maximiert Werte, ohne das Ziel wirklich zu treffen.

Verantwortung bleibt menschlich

Wer Modelle als Lösungen behandelt, gibt Verantwortung ab.

Beispiel: Ein statistisch starkes Modell kann in der Praxis falsche Entscheidungen stützen, wenn Kontext und Ziel nicht mitgedacht werden.

Probleme sind wichtiger als Algorithmen

Im Machine Learning zählt die Problemdefinition mehr als der Start mit dem Algorithmus.

Der häufigste Fehler im Machine Learning: Mit dem Algorithmus beginnen, nicht mit dem Problem.

1 Symptome vs. Ursachen

Symptome zu optimieren löst kein Problem. Ursachen zu verstehen braucht Domänenwissen und kritisches Fragen.

2 Unscharfe Fragen

Eine schlechte Problemdefinition führt zu falschen Antworten. Die Frage ist oft wichtiger als der Algorithmus.

3 Wenn ML nicht hilft

Oft sind einfachere Analysen, Prozessänderungen oder qualitative Methoden schneller und wirksamer.

📌 Problemverständnis ist die wichtigste Kompetenz im Machine Learning.

Die richtige Fragestellung

DER AUSGANGSPUNKT JEDES PROJEKTS

Die Qualität der Frage bestimmt die Qualität des Ergebnisses — in der Wissenschaft wie im Machine Learning.

1

Frageart klären

Klassifikation, Regression, Exploration oder Anomalieerkennung?

2

Fragen statt vorschneller Antworten

Wer zu früh antwortet, löst oft das falsche Problem.

3

Entscheidungskontext verstehen

Wer entscheidet? Wann? Unter welchen Bedingungen?

4

Von der Frage zum Modell

Der Übergang erfordert Reflexion — nicht nur Technik.

Beispiel: Ohne klare Frage wird das Modell präzise — aber für die falsche Aufgabe.

Zielvariablen bestimmen den Erfolg

Die Zielvariable ist wichtiger als der Algorithmus: Sie legt fest, *was* das Modell lernt — und welche Fehler akzeptabel sind.

Die unterschätzte Entscheidung

- Zielvariablen entstehen durch Entscheidungen, nicht objektiv
- Gute Zielvariablen sind messbar, relevant und schwer manipulierbar
- Sie müssen das eigentliche Ziel abbilden

Häufige Fehler bei der Wahl

- **Verwechslung von Ziel und Messgröße**
- **Verdeckte Zielkonflikte** bleiben unsichtbar
- **Proxy-Variablen** verzerren das Ergebnis

- ☐ Die Wahl der Zielvariable ist oft die folgenreichste Entscheidung im Projekt.

Goodharts Gesetz

„Wenn eine Kennzahl zum Ziel wird, hört sie auf, eine gute Kennzahl zu sein.“ — Charles Goodhart

Thesis: Sobald eine Kennzahl zum Ziel wird, verliert sie ihre Aussagekraft für das eigentliche Ziel.

Das Gesetz

Optimierung verschiebt den Fokus: Nicht das Phänomen zählt, sondern die Kennzahl.

Im Machine Learning

Accuracy kann seltene, kritische Klassen verdecken — trotz scheinbar guter Gesamtwerte.

In Management und Gesellschaft

KPIs, Tests und Ratings erzeugen Anreize, die das Metrikverhalten verzerren.

Die Schutzmaßnahme

Kennzahlen immer am eigentlichen Ziel prüfen — nie isoliert.

Konsequenz: 95 % Accuracy kann eine kritische Klasse komplett übersehen.

Trainingsdaten verstehen

DATEN ALS BEOBACHTUNGEN — NICHT ALS WAHRHEIT



Daten schlagen Modelle

Schlechte Daten erzeugen schlechte Modelle.



Repräsentativität ist entscheidend

Nur passende Daten tragen die Realität zuverlässig ab.



Daten zeigen Vergangenheit

Sie beschreiben, was war — nicht, was sein wird.



Mehr Daten sind nicht automatisch besser

Entscheidend ist der Entstehungskontext, nicht die Menge.

Daten sind nicht neutral

BIAS UND VERZERRUNGEN

Die Verzerrung steckt oft schon in den Daten.

Typische Quellen

- **Historische Ungleichheiten:** Daten reproduzieren vergangene Ungleichheiten
- **Auswahlverzerrungen:** Wer erhoben wird, bestimmt das Ergebnis
- **Messfehler:** Systematische Fehler gehen direkt ins Modell ein
- **Proxy-Variablen:** Stellvertreter können unbeabsichtigt diskriminieren

Konsequenz: Wer erst beim Modell auf Fairness prüft, greift zu spät an.

Fairness beginnt bei den Daten

Verzerrte Trainingsdaten liefern verzerrte Modelle.

Fairness muss bei Beschaffung und Analyse beginnen — nicht erst im Modell.

- ❏ Systematisch verzerrte Daten lassen sich nicht vollständig „wegmodellieren“.

Features als Hypothesen

Features spiegeln Annahmen über die Wirklichkeit wider — nicht objektive Fakten.

Features modellieren Wirklichkeit

Feature Engineering bringt Problemverständnis ins Modell.
Features sind Hypothesen — und können falsch sein.

Mehr ist nicht besser

Mehr Features erhöhen Dimensionalität und Overfitting-Risiko.
Relevant ist wichtiger als zahlreich.

Proxy-Features sind riskant

Stellvertreter können schützende Merkmale indirekt abbilden —
und dennoch diskriminierende Modelle erzeugen.

Domänenwissen bleibt zentral

Gute Feature-Auswahl braucht Fachverständnis. Technische
Kompetenz allein reicht nicht.

Baselines als Realitätstest

SCHUTZ VOR SELBSTTÄUSCHUNG

Eine gute Modellbewertung braucht immer eine Baseline.

Was eine Baseline ist

Das einfachste sinnvolle Vergleichsmodell: Was leistet das komplexe Modell zusätzlich?

- Häufigste Klasse vorhersagen
- Mittelwert vorhersagen
- Experten als Referenz

Konsequenz: Ohne Baseline bleibt jede Kennzahl schwer einzuordnen.

Warum sie unverzichtbar ist

- Ohne Baseline fehlt der Maßstab
- Einfache Modelle können gewinnen
- Komplexität ist kein Selbstzweck

Modelle als Denkwerkzeuge

Modelltypen stellen unterschiedliche Fragen. Die Wahl ist konzeptuell, nicht technisch.



Klassifikation

Zuordnung zu Kategorien.



Regression

Schätzung kontinuierlicher Werte.



Clustering

Verborgene Gruppen ohne Vorgaben entdecken.



Anomalieerkennung

Auffälligkeiten und Abweichungen erkennen.

Der typische Machine-Learning-Workflow

Der Workflow ist wichtiger als der Algorithmus. Wer ihn versteht, trifft bessere Entscheidungen in jedem Schritt.



Beispiel: Neue Erkenntnisse aus der Bewertung führen oft zurück zu Daten oder Merkmalen. Der Prozess ist bewusst iterativ.

Praxisbeispiel: Kreditwürdigkeitsprüfung

Ein typisches Anwendungsfeld für binäre Klassifikation: Kredit gewähren oder ablehnen.



Das Ziel

Kreditwürdigkeit vorhersagen: kreditwürdig oder nicht kreditwürdig.



Die Eingabe

Wichtige Merkmale: Einkommen, Beschäftigungsdauer, Verbindlichkeiten, Zahlungsverhalten, Bonität.



Das Modell

Lernt aus früheren Anträgen, welche Muster auf Ausfallrisiko hindeuten.



Die Ausgabe

Eine Wahrscheinlichkeit plus klare Empfehlung: gewähren oder ablehnen.

Konsequenz: Die Entscheidung wird schneller, konsistenter und besser begründbar.

Praxisbeispiel: Entscheidungsbaum vs. Random Forest

MODELLVERGLEICH

Ein Vergleich zweier Modelltypen zeigt den zentralen Trade-off zwischen **Interpretierbarkeit** und **Leistungsfähigkeit**.

Entscheidungsbaum

Intuitiv nachvollziehbar, aber anfällig für Überanpassung.

- **Interpretierbarkeit:** Sehr hoch.
- **Overfitting-Risiko:** Hoch.
- **Genauigkeit:** Häufig geringer.

Random Forest

Robuster durch die Kombination vieler Bäume, dafür weniger transparent.

- **Interpretierbarkeit:** Gering.
- **Overfitting-Resistenz:** Deutlich höher.
- **Genauigkeit:** Meist höher und stabiler.

☐ Je höher der Bedarf an Erklärbarkeit, desto eher spricht man für den Entscheidungsbaum — je höher der Performance-Anspruch, desto eher für den Random Forest.

Praxisbeispiel: Kreditentscheidung – Human-in-the-Loop

Bei sensiblen Entscheidungen ergänzt menschliches Urteilsvermögen die Effizienz von KI. So entstehen Verantwortung, Kontext und Tempo.

Modell: Daten & Fakten

Berechnet Ausfallrisiken und markiert relevante Signale.

Mensch: Kontext & Werte

Prüft Lebensumstände, Marktkenntnis und Ethik.

Stärke: Effizienz & Ethik

Schneller Prozess, weniger Fehlentscheidungen, mehr Fairness.

Konsequenz: Die finale Kreditentscheidung bleibt nachvollziehbar und verantwortbar.

Praxisbeispiel: Human-in-the-Loop – Das Prinzip

Human-in-the-Loop (HitL) integriert menschliche Expertise in KI-Entscheidungen, damit Effizienz, Verantwortung und Kontext zusammenwirken.



Medizin

KI unterstützt die Diagnostik; Ärzte entscheiden final unter Einbezug von Präferenzen und Ethik.



Justiz

Algorithmen bewerten Risiken; Richter berücksichtigen individuelle Umstände und Recht.



Finanzen

KI erkennt Muster; Berater geben passende Empfehlungen nach Ziel und Risikoprofil.



Personalwesen (HR)

Modelle filtern vor; Personalmanager führen Gespräche und treffen die Einstellungsentscheidung.

HitL ist ein Zeichen von Reife: KI bleibt effektiv, der Mensch bleibt verantwortlich.

Warum Kennzahlen allein nicht reichen

Kennzahlen messen — sie bewerten nicht.

Kennzahl ≠ Urteil

Eine Kennzahl zeigt nur, was gemessen wurde. Ob es gut genug ist, braucht Kontext und Ziele.

Gute Werte, falsche Entscheidung

Ein Modell kann starke Kennzahlen haben und in der Praxis trotzdem scheitern.

Bewertung ist kontextabhängig

Welche Kennzahl zählt, ist eine Wertentscheidung — nicht nur eine technische Frage.

Kontinuierliche Prüfung

Verantwortbare KI braucht fortlaufende, kritische Bewertung statt einmaliger Messung.

Accuracy kann täuschen

Das Problem mit Accuracy

- Beliebt — aber bei unausgeglichene Klassen oft irreführend
- **95 % Accuracy** ist möglich, ohne einen positiven Fall zu erkennen
- Sie misst nur Trefferquoten, nicht Fehlerarten oder Kosten
- Goodharts Gesetz: Das Ziel wird zur Falle

Ein konkretes Beispiel

Bei 95 % negativen Fällen erreicht ein Modell mit immer „negativ“ **95 % Accuracy** — und erkennt keinen einzigen positiven Fall.

Exzellente auf dem Papier, wertlos in der Praxis.

Accuracy

Einfach, aber verzerrungsanfällig.

Precision

Wie viele positive Vorhersagen stimmen.

Recall

Wie viele positive Fälle gefunden werden.

F1-Score

Der Ausgleich zwischen Precision und Recall.

Precision und Recall als Trade-off

Welche Fehlerart teurer ist, hängt vom Kontext ab — nicht vom Modell.

Die Kennzahlen

- **Precision:** Wie viele positive Vorhersagen sind korrekt?
- **Recall:** Wie viele positive Fälle wurden erkannt?
- **F1-Score:** Kompromiss aus beiden — aber kein Ersatz für Urteilskraft.

Der Trade-off

Mehr Precision heißt oft weniger Recall — und umgekehrt.

- **Medizin:** Recall ist kritisch — kein positiver Fall darf übersehen werden.
- **Spam-Filter:** Precision ist kritisch — wichtige E-Mails dürfen nicht blockiert werden.
- **Kreditvergabe:** Beide Fehlerarten haben reale Folgen.

Bewertung & Modellwahl — beides sind Entscheidungen

Es gibt keine objektive Bewertung und kein universell bestes Modell — beides erfordert Urteilskraft.

Bewertung ist eine Wertentscheidung

01

Welche Kennzahl zählt?

Wertentscheidung, keine Technikfrage.

02

Welcher Fehler ist teurer?

Kosten hängen vom Kontext ab.

03

Schwellenwerte steuern den Trade-off

Bewusst setzen, nicht automatisch.

04

Bewertung schützt vor Selbsttäuschung

Die Illusion des besten Modells

Mehr als Genauigkeit

Transparenz, Fairness, Kosten, Wartbarkeit.

Modellwahl braucht Urteilskraft

Modelle sind Alternativen, keine Siegerlisten.

Trade-offs verstehen

Jede Modellentscheidung ist ein Trade-off — keine kostenlose Optimierung.

Genauigkeit vs. Transparenz

Mehr Genauigkeit bedeutet oft weniger Interpretierbarkeit.



Genauigkeit vs. Fairness

Maximierung der Gesamtgenauigkeit kann Gruppen benachteiligen.



Geschwindigkeit vs. Qualität

Tempo spart jetzt Zeit, erzeugt später oft technische Schulden.

Risiken bewerten

Risiko entsteht nicht durch Fehler allein, sondern durch ihre Folgen.

Datenrisiken

Bias, fehlende Werte, Messfehler, Repräsentativitätslücken.

Modellrisiken

Overfitting, Underfitting, Datenlecks, Drift.

Entscheidungsrisiken

False Positives und False Negatives mit realen Folgen.

Risikobewertung als Haltung

Die Frage ist nicht nur: Wie wahrscheinlich ist der Fehler? Sondern: Was passiert dann — und für wen?

Beispiel: Ein seltenes Fehurteil kann für Betroffene gravierender sein als ein häufiger, aber harmloser Fehler.

Unsicherheit kommunizieren

Unsicherheit offen zu benennen schafft Glaubwürdigkeit — nicht Schwäche.

Warum sie verschwindet

- Stakeholder wollen klare Antworten, keine Bandbreiten
- So entsteht Scheingewissheit statt Transparenz
- Ehrlichkeit über Unsicherheit baut Vertrauen auf

Werkzeuge

- **Wahrscheinlichkeiten** mit Kontext kommunizieren
- **Konfidenzintervalle** sichtbar machen
- **Szenarien statt Prognosen**
- **Sensitivitätsanalysen** für Annahmen nutzen

📄 Ehrlichkeit über Unsicherheit ist ein Zeichen von Kompetenz — nicht von Schwäche.

Von Vorhersagen zu Handlungsempfehlungen

Empfehlungen enthalten Wertentscheidungen. Der Übergang von Vorhersage zu Empfehlung muss bewusst und nachvollziehbar gestaltet werden.

1

Vorhersage

Modellbasierte Wahrscheinlichkeitsaussage

2

Interpretation

Kontext klären, Alternativen sichtbar machen

3

Begründung

Argumentation mit Daten, Warrant und Grenzen

4

Empfehlung

Handlung unter Unsicherheit – transparent und prüfbar

Verantwortbare Entscheidungen

Modelle liefern Orientierung. Verantwortung bleibt menschlich.

Modelle unterstützen

Wahrscheinlichkeiten und Muster —
kraftvolle Werkzeuge ohne Werte.

Menschen entscheiden

Kontext, Werte und Urteilskraft fließen in
die Entscheidung ein.

Menschen verantworten

Verantwortung braucht Intentionalität und
Rechenschaftspflicht.

Algorithmen sind nicht neutral

Sie tragen die Wertentscheidungen von Daten, Zielen und Entwicklern in
sich.

Verantwortung als Haltung

Sie beginnt bei der Datenerhebung und endet erst mit der
Entscheidung.

Was Leser aus dem Buch mitnehmen

Das Buch schärft den Blick für bessere Modellarbeit — von der Kritik am Modell bis zur verantwortungsvollen Entscheidung.



Kritisches Denken

Modelle und Kennzahlen hinterfragen. Die richtigen Fragen vor, während und nach der Modellierung stellen.



Bessere Modellbewertung

Über Accuracy hinaus denken: Precision, Recall, Fairness, Transparenz und Kontextsensitivität.



Bessere Entscheidungen

Vorhersagen in begründete Handlungsempfehlungen übersetzen — bewusst und strukturiert.



Umgang mit Unsicherheit

Wahrscheinlichkeiten korrekt kommunizieren. Szenarien statt falscher Gewissheiten.



Verantwortungsvolle KI

Bias, Fairness und Transparenz als praktische Anforderungen im Alltag.

Beispiel: Nicht die höchste Accuracy zählt, sondern ein Modell, das im Kontext tragfähig und fair entscheidet.

Fazit

Machine Learning liefert Vorhersagen — Entscheidungen bleiben menschlich.



„Die eigentliche Herausforderung des Machine Learning beginnt nicht beim Modell — sie beginnt dort, wo Entscheidungen getroffen werden.“

Das Buch plädiert für kritisches Denken, methodische Sorgfalt und menschliche Verantwortung — von den Daten bis zur Entscheidung.

Quellen und theoretische Grundlagen I

Denken, Urteilen und Entscheiden

Ellmann, Mathias: *Von Modellen zu Entscheidungen in Machine Learning mit Python. Vom Modell zur verantwortbaren Entscheidung*. BookRix, München, 2026. ISBN: 978-3-6952-6182-6. Grundlage der Präsentation und zentrale Referenz für die These, dass Machine Learning nicht beim Modell endet, sondern bei der verantwortbaren Entscheidung.

Ellmann, Mathias: *Kritisches Denken in Data Science mit Python. Vom Beobachten zur verantwortbaren Entscheidung*. BookRix, München, 2026. ISBN: 978-3-6952-6078-2. Grundlage für die Verbindung von datenbasierter Analyse, kritischem Denken, Interpretation, Urteilkraft und verantwortbarer Entscheidung.

Ellmann, Mathias: *Von Daten zu Lösungen in Data Science mit Python. Vom Problem zur verantwortbaren Lösung*. BookRix, München, 2026. ISBN: 978-3-6952-6121-5. Kontextquelle für die übergeordnete Perspektive, dass datenbasierte Arbeit nicht bei Analyse oder Modellierung endet, sondern auf verantwortbare Lösungen für reale Probleme ausgerichtet ist.

Kahneman, Daniel: *Schnelles Denken, langsames Denken*. Siedler Verlag, München, 2012. Grundlage für die Unterscheidung zwischen intuitivem und reflektiertem Denken sowie für typische Verzerrungen bei Urteilen, Prognosen und Entscheidungen.

Tversky, Amos; Kahneman, Daniel: „Judgment under Uncertainty. Heuristics and Biases.“ In: *Science* 185(4157), 1974, S. 1124–1131. Grundlage für die Darstellung von Heuristiken, Denkfehlern und Urteilen unter Unsicherheit.

Korzybski, Alfred: *Science and Sanity. An Introduction to Non-Aristotelian Systems and General Semantics*. Institute of General Semantics, Lakeville, Connecticut, 1958. Grundlage für die Unterscheidung zwischen Wirklichkeit und Beschreibung: Modelle, Daten und Diagramme sind Landkarten der Wirklichkeit, nicht die Wirklichkeit selbst.

Hammond, John S.; Keeney, Ralph L.; Raiffa, Howard: *Smart Choices. A Practical Guide to Making Better Decisions*. Harvard Business Review Press, Boston, 2015. Grundlage für die entscheidungsanalytische Perspektive, nach der gute Entscheidungen klare Ziele, Alternativen, Abwägungen und Konsequenzen benötigen.

Quellen und theoretische Grundlagen II

Machine Learning, Modelle und Bewertung

Burkov, Andriy: *The Hundred-Page Machine Learning Book*. Andriy Burkov, 2019. Grundlage für eine kompakte und systematische Darstellung zentraler Machine-Learning-Konzepte, Modelltypen und Bewertungsfragen.

Géron, Aurélien: *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 3. Auflage. O'Reilly Media, Sebastopol, 2022. Grundlage für praxisnahe Machine-Learning-Workflows mit Python, Scikit-Learn und modernen Modellierungsverfahren.

Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome: *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. 2. Auflage. Springer, New York, 2009. Grundlage für das statistische Verständnis von Lernen, Modellierung, Generalisierung, Overfitting, Underfitting und Vorhersage.

Provost, Foster; Fawcett, Tom: *Data Science for Business*. O'Reilly Media, Sebastopol, 2013. Grundlage für die Verbindung von datenbasierten Modellen, geschäftlichen Fragestellungen, Modellbewertung und praktischer Entscheidungsunterstützung.

Goodhart, C. A. E.: „Problems of Monetary Management: The UK Experience.“ In: *Monetary Theory and Practice: The UK Experience*. Macmillan Education UK, London, 1984, S. 91–121. Grundlage für Goodharts Gesetz und die Aussage, dass Kennzahlen ihre Aussagekraft verlieren können, sobald sie selbst zum Ziel werden.

Saito, Takaya; Rehmsmeier, Marc: „The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets.“ In: *PLOS ONE* 10(3), 2015, e0118432. Grundlage für die Darstellung von Precision, Recall und F1-Score bei unausgeglichenen Klassifikationsproblemen.

Quellen und theoretische Grundlagen III

Fairness, Erklärbarkeit und verantwortbare KI

Barocas, Solon; Hardt, Moritz; Narayanan, Arvind: *Fairness and Machine Learning. Limitations and Opportunities*. MIT Press, Cambridge, MA, 2023. Grundlage für die Auseinandersetzung mit Fairness, Bias, Diskriminierungsrisiken und den Grenzen technischer Fairness-Kriterien im Machine Learning.

Floridi, Luciano u. a.: „AI4People—An Ethical Framework for a Good AI Society. Opportunities, Risks, Principles, and Recommendations.“ In: *Minds and Machines* 28, 2018, S. 689–707. Grundlage für die ethische Einordnung von KI-Systemen und für Prinzipien wie Transparenz, Verantwortung, Nichtschädigung und menschliche Kontrolle.

O’Neil, Cathy: *Weapons of Math Destruction*. Crown, New York, 2016. Grundlage für die kritische Betrachtung algorithmischer Entscheidungssysteme, insbesondere ihrer gesellschaftlichen Folgen, Skalierungseffekte und potenziellen Intransparenz.

Molnar, Christoph: *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. 2022. Grundlage für die Darstellung interpretierbarer Modelle, Modellgrenzen und Erklärbarkeit als Voraussetzung verantwortbarer Entscheidungen.

Ribeiro, Marco Tulio; Singh, Sameer; Guestrin, Carlos: „Why Should I Trust You? Explaining the Predictions of Any Classifier.“ In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, S. 1135–1144. Grundlage für die Bedeutung lokaler Modellerklärungen und für die Frage, warum Nutzer einzelnen Vorhersagen vertrauen sollten.

Sculley, D. u. a.: „Hidden Technical Debt in Machine Learning Systems.“ In: *Proceedings of the 29th International Conference on Neural Information Processing Systems*, Volume 2, MIT Press, 2015, S. 2503–2511. Grundlage für die Betrachtung produktiver Machine-Learning-Systeme, technischer Schulden, Wartbarkeit und langfristiger Betriebsrisiken.