**Buchbezogene Merkmale**

| **Feature** | **Beschreibung** |
| --- | --- |
| **Titel** | (optional, nur für NLP-Modelle) – kann Hinweise auf Genre/Thema geben |
| **Genre** | z. B. Thriller, Romanze, Fantasy – visuell starke Genres werden häufiger verfilmt |
| **Zielgruppe** | Kinder, Jugendliche, Erwachsene – beeinflusst Vermarktungschancen |
| **Sprache** | Originalsprache – englischsprachige Bücher haben größere Märkte |
| **Erscheinungsjahr** | Zeitlicher Kontext – neue Bücher haben aktuellere Chancen, Klassiker einen anderen Status |
| **Seitenanzahl** | Extrem lange oder kurze Bücher könnten Einfluss auf Verfilmbarkeit haben |

**🧑‍💼 Autorenbezogene Merkmale**

| **Feature** | **Beschreibung** |
| --- | --- |
| **Bekanntheitsgrad** | Anzahl veröffentlichter Bücher, Bestseller-Status |
| **Vergangene Verfilmungen** | Wurde der Autor schon verfilmt? Erfolgreich? |
| **Social-Media-Präsenz** | Fanbase, Reichweite |

**📈 Kommerzielle Merkmale**

| **Feature** | **Beschreibung** |
| --- | --- |
| **Verkaufszahlen** | Hohe Verkaufszahlen korrelieren oft mit Verfilmungen |
| **Ranking in Bestsellerlisten** | z. B. New York Times, Spiegel etc. |
| **Anzahl Übersetzungen** | Internationale Reichweite erhöht Chancen auf Verfilmung |
| **Piraterie-Rate** | (falls vorhanden) – kann auf Popularität hindeuten |

**🌟 Bewertungen & Rezeption**

| **Feature** | **Beschreibung** |
| --- | --- |
| **Durchschnittliche Bewertung** | z. B. auf Goodreads, Amazon |
| **Anzahl Bewertungen/Rezensionen** | Aktivität & Beliebtheit des Buchs |
| **Literarische Auszeichnungen** | z. B. Pulitzer, Booker Prize – erhöht mediales Interesse |

**🎬 Verlags- & Industrie-Daten (wenn verfügbar)**

| **Feature** | **Beschreibung** |
| --- | --- |
| **Verlagstyp** | Großer Verlag = bessere Verbindungen zur Filmbranche |
| **Optionierung durch Filmstudio** | Falls du diese Info bekommst, ist sie direkt relevant |
| **Genre-Verfilmungstrends** | Historisch höhere Verfilmungsrate bestimmter Genres (Fantasy, Thriller, Sci-Fi etc.) |

**💡 Optional: Textbasierte Features (NLP)**

Wenn du die **Beschreibung oder den Text** eines Buchs hast, kannst du:

* **Themen und Stimmungen (Sentiment)** extrahieren
* **Named Entities (Personen, Orte etc.)** analysieren
* **Narrative Struktur oder Dialoganteil** bewerten

Diese NLP-Features kannst du v. a. mit größeren Datensätzen (z. B. 1000+ Bücher) nutzen.

## 🧠 1. **Modelltraining vs. Vorhersage für neue Bücher**

* Du **trainierst** dein Modell mit Daten von **älteren Büchern**, bei denen du **weißt, ob sie verfilmt wurden oder nicht** (mit deinem aktuellen Datensatz).
* Danach willst du das Modell benutzen, um für ein **neues Buch**, das noch **nicht verfilmt wurde**, vorherzusagen:  
  → **Wie hoch ist die Wahrscheinlichkeit, dass es verfilmt wird?**

Das ist ein typischer **Supervised Learning Task** (binäre Klassifikation).

## ✅ 2. **Deine Features sind geeignet für neue Bücher**

Wichtig ist: Du brauchst nur Features, die du **auch bei einem neuen Buch kennst**.  
Und das ist bei dir der Fall:

| **Feature** | **Auch bei neuem Buch verfügbar?** | **Gut für Prediction?** |
| --- | --- | --- |
| Titel | ✅ | (optional) |
| Genre | ✅ | ✅ |
| Zielgruppe | ✅ | ✅ |
| Sprache | ✅ | ✅ |
| Jahr | ✅ | ✅ (neu → aktueller Markt) |
| Autor-Bekanntheit | ✅ (z. B. Anzahl vorheriger Bücher, Follower) | ✅ |
| Verkaufszahlen | ⚠️ Nur teilweise verfügbar zu Beginn | Sehr wichtig |
| Durchschnittliche Bewertung | ⚠️ Nicht sofort verfügbar | Nützlich |
| Anzahl Bewertungen | ⚠️ Wächst mit Zeit | Nützlich |
| Verlag | ✅ | ✅ |

### 🔍 Empfehlung für neue Bücher:

Wenn du **ein neues Buch gleich zur Veröffentlichung** bewerten willst, dann:

* Trainiere dein Modell auch auf Büchern **kurz nach deren Veröffentlichung**, mit den damals bekannten Features (z. B. erste Bewertungen, Verkaufsrang etc.)
* Alternativ: Trainiere zwei Modelle:
  + Eins für Bücher **direkt beim Erscheinen** (ohne spätere Metriken)
  + Eins für Bücher **nach 6 Monaten**, wenn mehr Verkaufsdaten vorliegen

## 🧰 Tools & Modellideen

* Klassische Modelle: Logistic Regression, Random Forest, XGBoost
* Wahrscheinlichkeitsausgabe: statt nur 0 oder 1 → z. B. „Verfilmung mit 82 % Wahrscheinlichkeit“
* Wichtig: **Train/Test-Split nach Jahr** – nicht zufällig – damit du nicht Zukunftsdaten für die Vergangenheit verwendest