

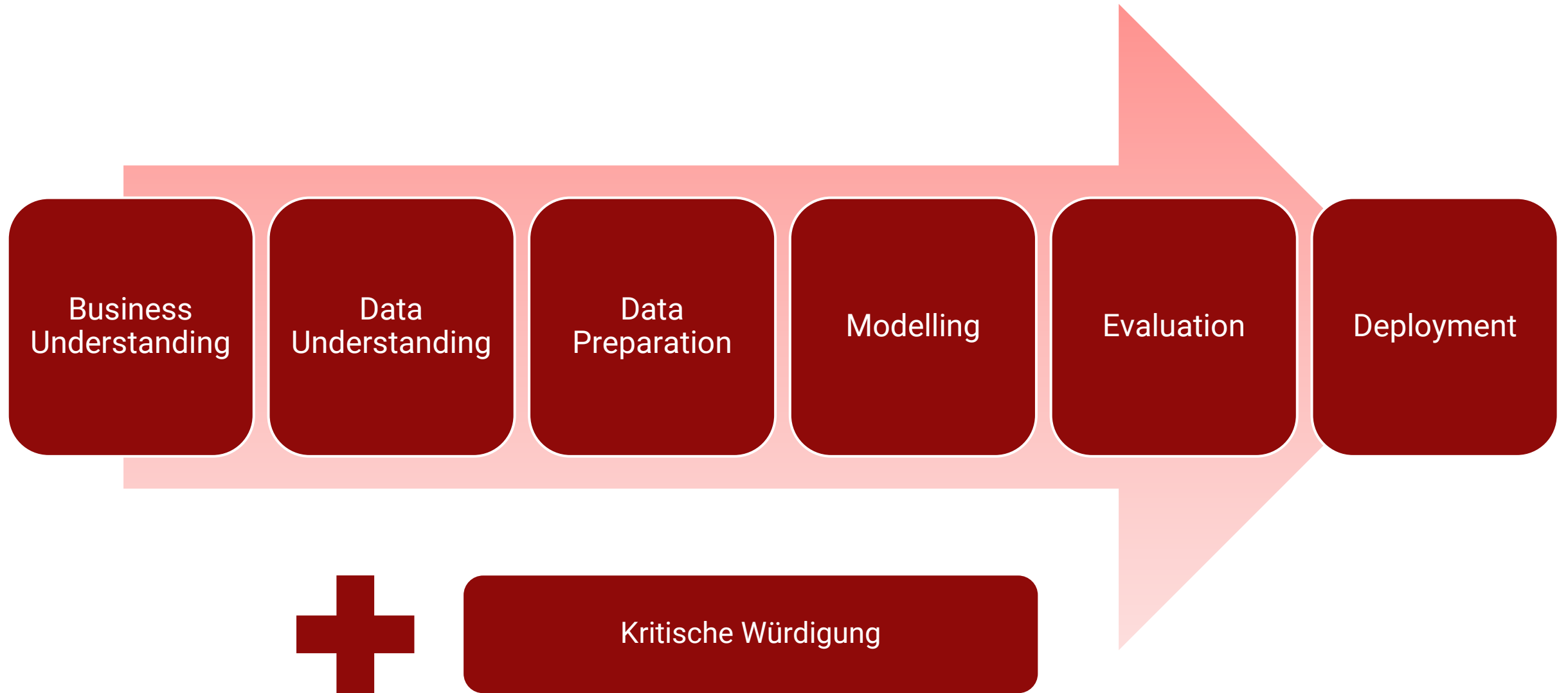
KI-Projekt Filmvorschläge

Marcel Bulling, Johanna Deike, Aidan Zimmer, Lisa Reiß-Park, Luca Chmielarski & Nora Klemp

Public



Agenda nach CRISP-DM



Use Cases - Kunden

- Personalisierte Empfehlungen für bessere Auswahl
- Präferenz für das Kino, Zeitersparnis bei Buchung → Kundenbindung
- Wird auf Filme aufmerksam und entschließt sich dann diese zu schauen

Höhere Kundenzufriedenheit &
Bindung an das Kino

Business
Understanding

Data
Understanding

Data
Preparation

Modelling

Evaluation

Deployment



Use Cases - Kinobesitzer

- Verbesserter Kundenservice
- Genauere Planung von Personal, Sälen und Verpflegung
- Höheres Potential für Cross- und Upselling
- Effizienteres Marketing

Höherer Umsatz &
Wettbewerbsvorteile

Business
Understanding

Data
Understanding

Data
Preparation

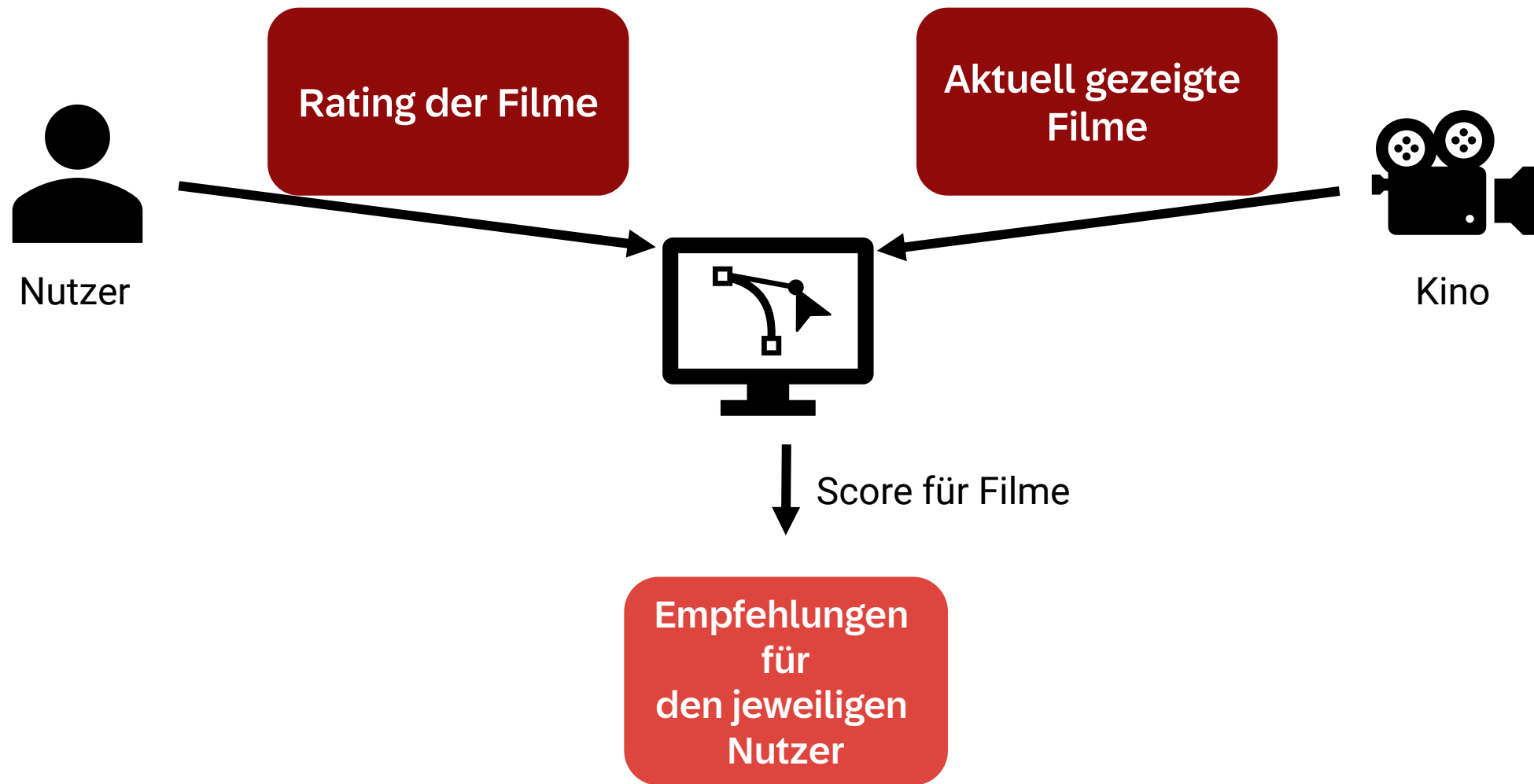
Modelling

Evaluation

Deployment



Überblick bereitgestellter Endpunkt



Business
Understanding

Data
Understanding

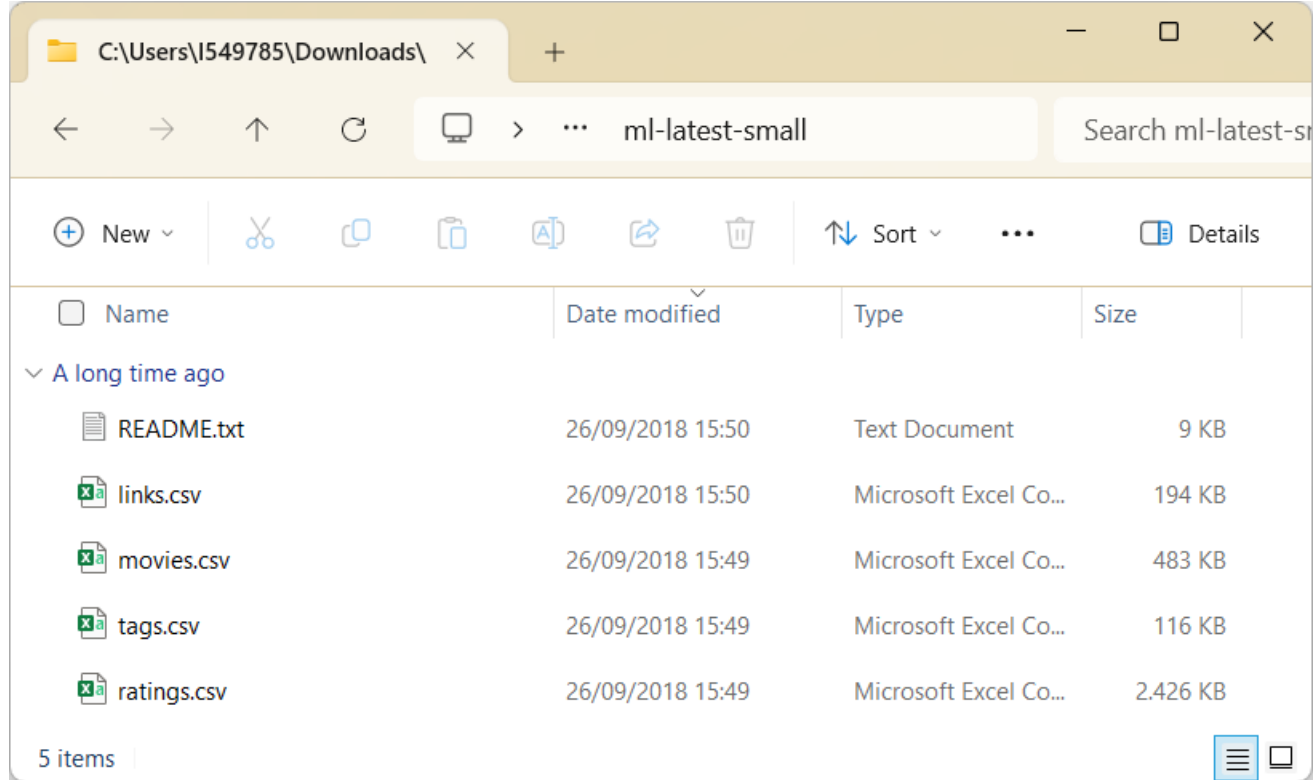
Data
Preparation

Modelling

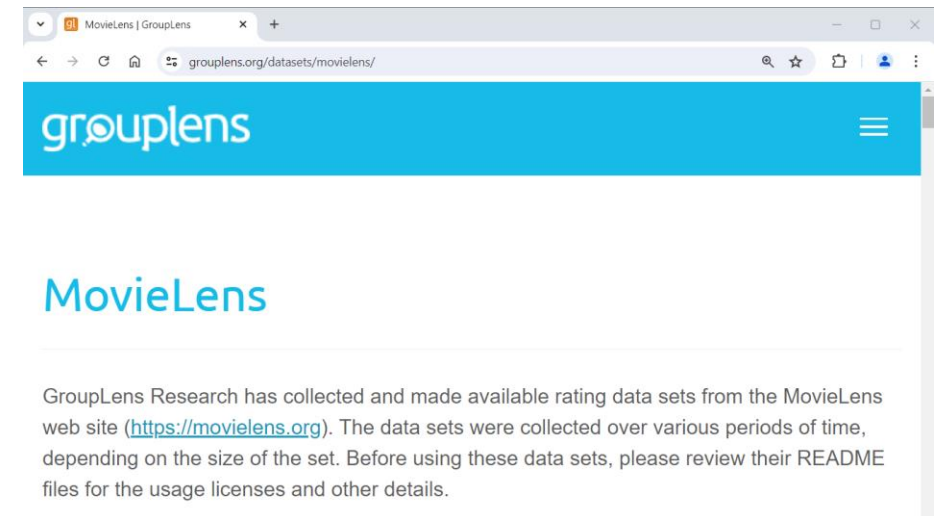
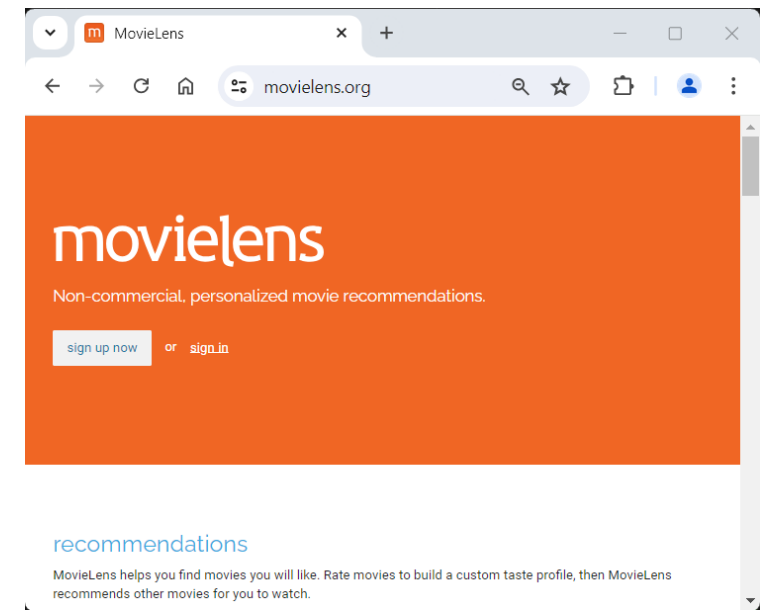
Evaluation

Deployment

Data Understanding



Name	Date modified	Type	Size
▼ A long time ago			
README.txt	26/09/2018 15:50	Text Document	9 KB
links.csv	26/09/2018 15:50	Microsoft Excel Co...	194 KB
movies.csv	26/09/2018 15:49	Microsoft Excel Co...	483 KB
tags.csv	26/09/2018 15:49	Microsoft Excel Co...	116 KB
ratings.csv	26/09/2018 15:49	Microsoft Excel Co...	2.426 KB



Business
Understanding

Data
Understanding

Data
Preparation

Modelling

Evaluation

Deployment

Data Understanding

movies.csv

	A	B	C
1	movieId	title	genres
2	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy
3	2	Jumanji (1995)	Adventure Children Fantasy
4	3	Grumpier Old Men (1995)	Comedy Romance
5	4	Waiting to Exhale (1995)	Comedy Drama Romance
6	5	Father of the Bride Part II (1995)	Comedy
7	6	Heat (1995)	Action Crime Thriller
8	7	Sabrina (1995)	Comedy Romance
9	8	Tom and Huck (1995)	Adventure Children
10	9	Sudden Death (1995)	Action
11	10	GoldenEye (1995)	Action Adventure Thriller
12	11	American President, The (1995)	Comedy Drama Romance
13	12	Dracula: Dead and Loving It (1995)	Comedy Horror

ratings.csv

	A	B	C	D
1	userId	movieId	rating	timestamp
2	1	1	4,0	964982703
3	5	1	4,0	847434962
4	7	1	4,5	1106635946
5	15	1	2,5	1510577970
6	17	1	4,5	1305696483
7	18	1	3,5	1455209816
8	19	1	4,0	965705637
9	21	1	3,5	1407618878
10	27	1	3,0	962685262
11	31	1	5,0	850466616
12	32	1	3,0	856736119
13	33	1	3,0	939647444
14	40	1	5,0	832058959
15	43	1	5,0	848993983
16	44	1	3,0	869251860
17	45	1	4,0	951170182
18	46	1	5,0	834787906
19	50	1	3,0	1514238116
20	54	1	3,0	830247330

Business
Understanding

Data
Understanding

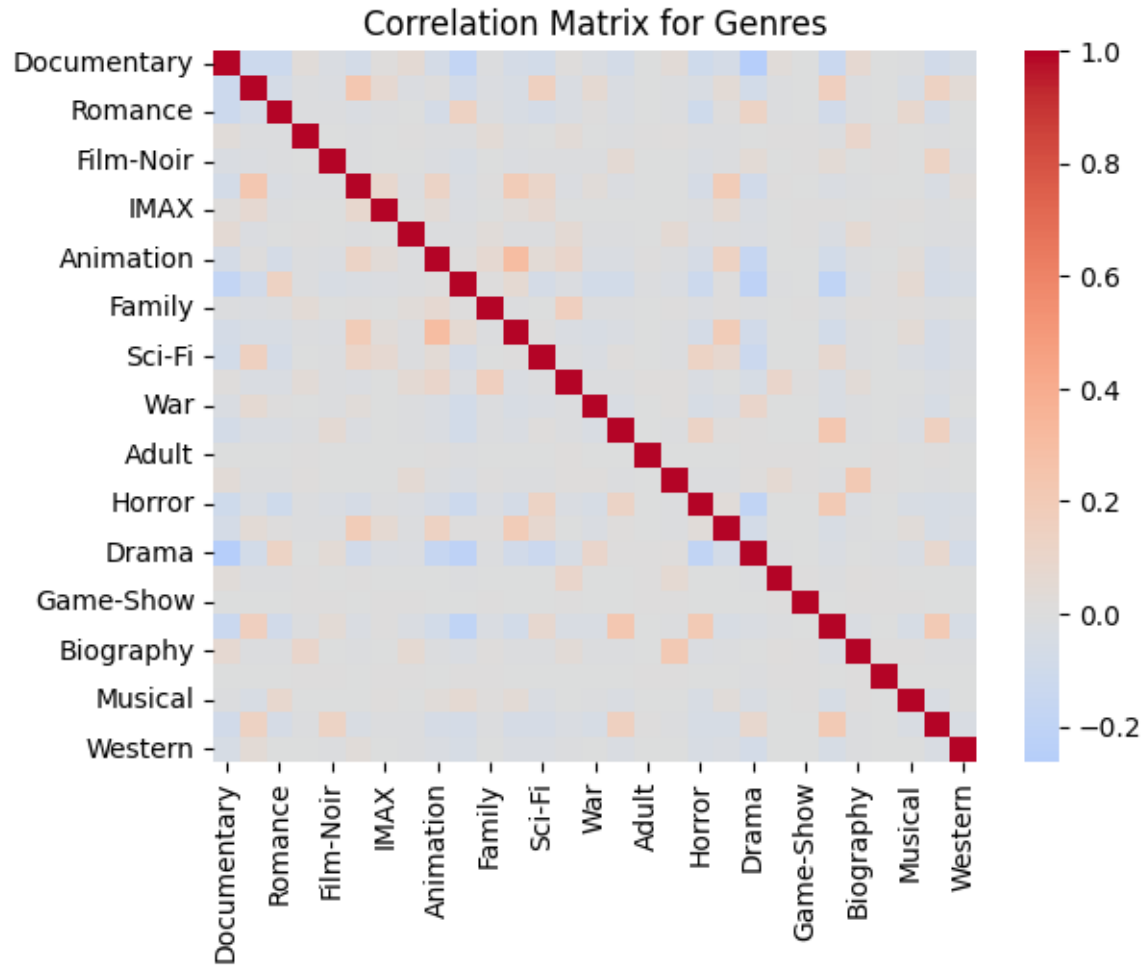
Data
Preparation

Modelling

Evaluation

Deployment

Data Understanding



Business
Understanding

Data
Understanding

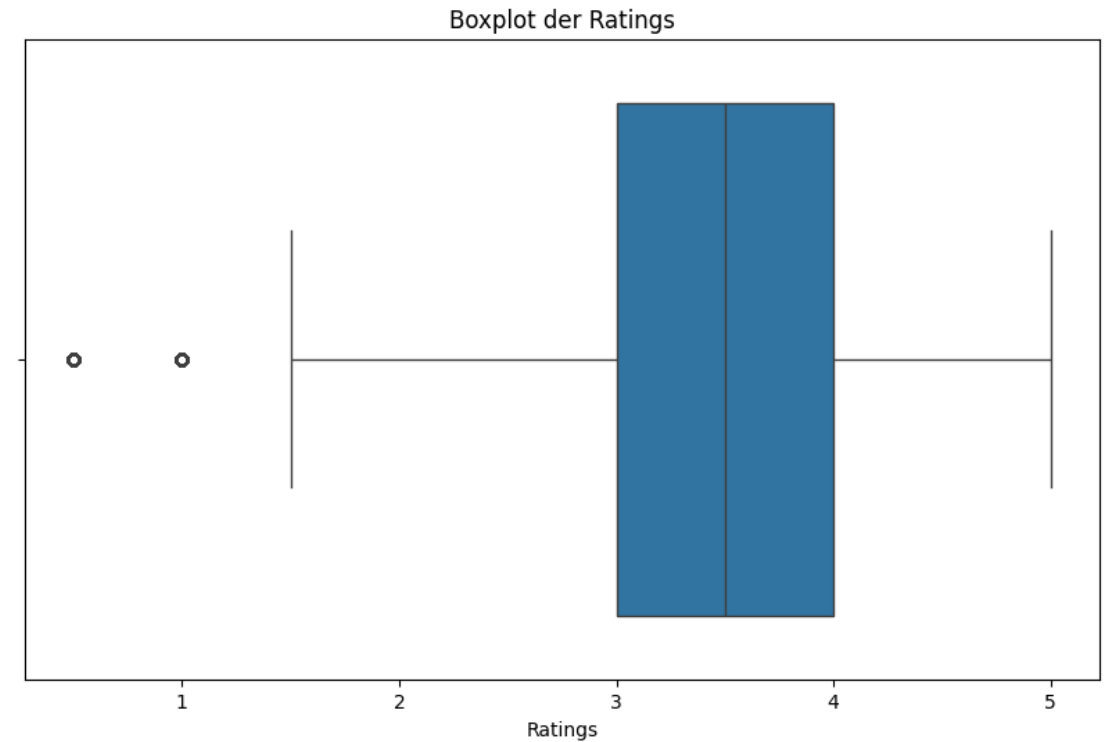
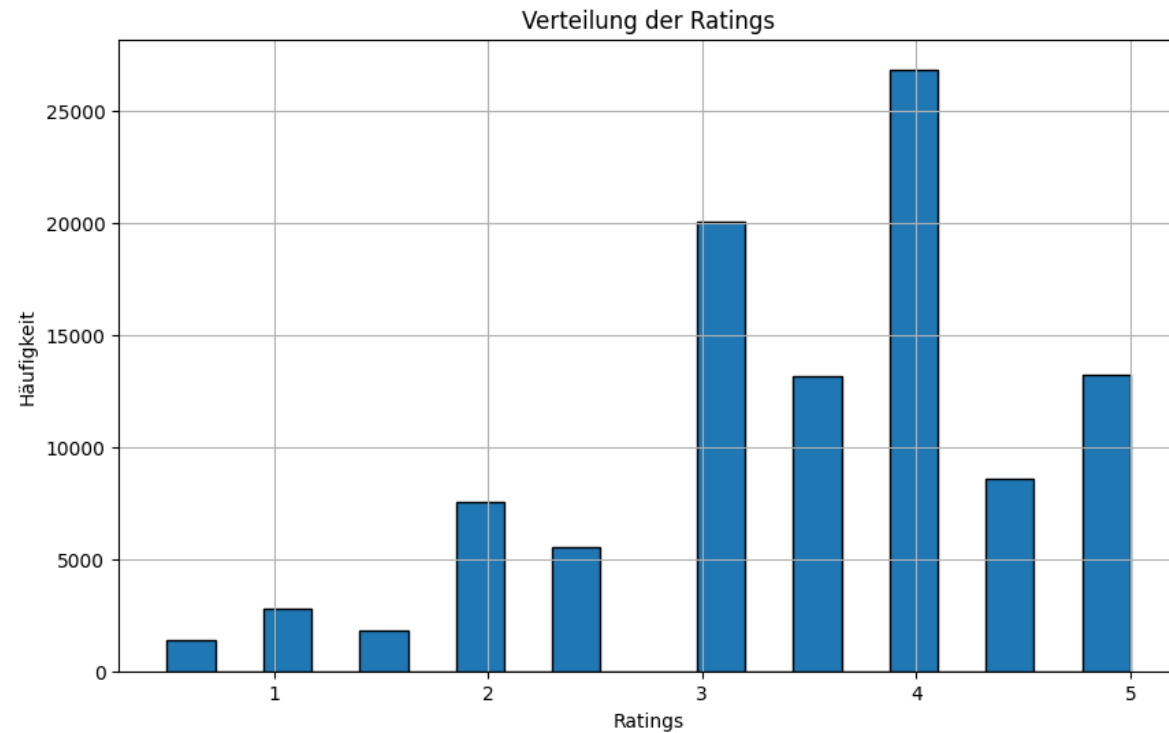
Data
Preparation

Modelling

Evaluation

Deployment

Data Understanding



Business
Understanding

Data
Understanding

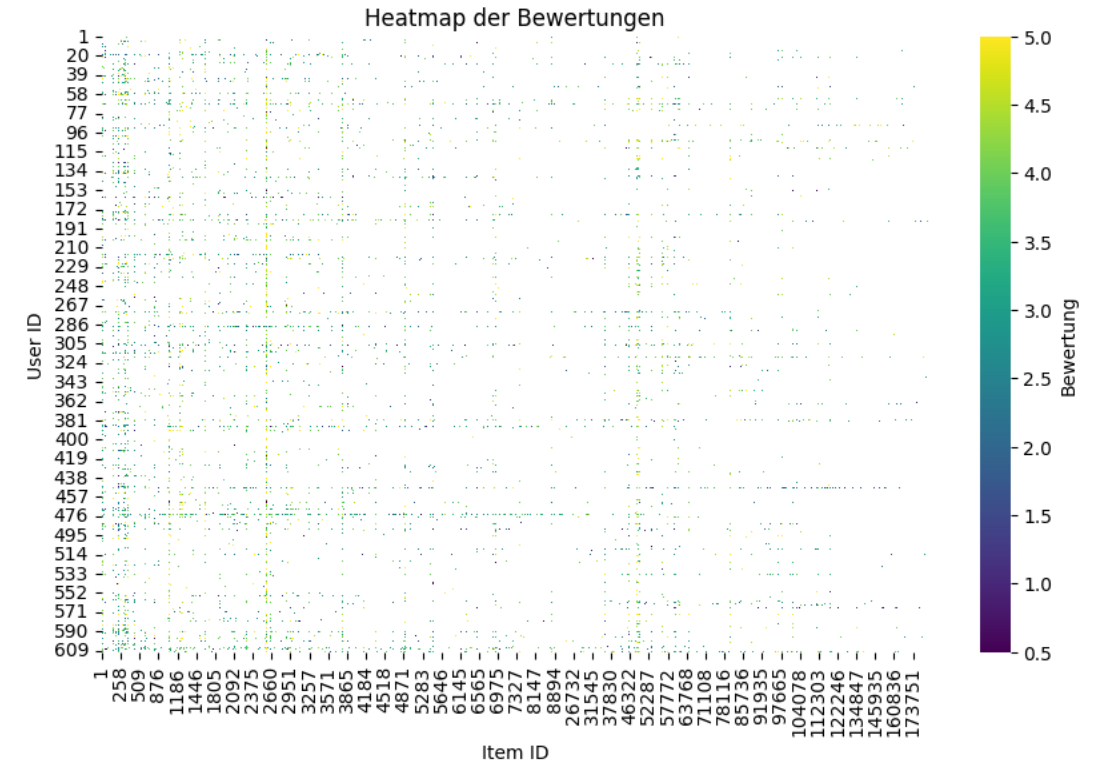
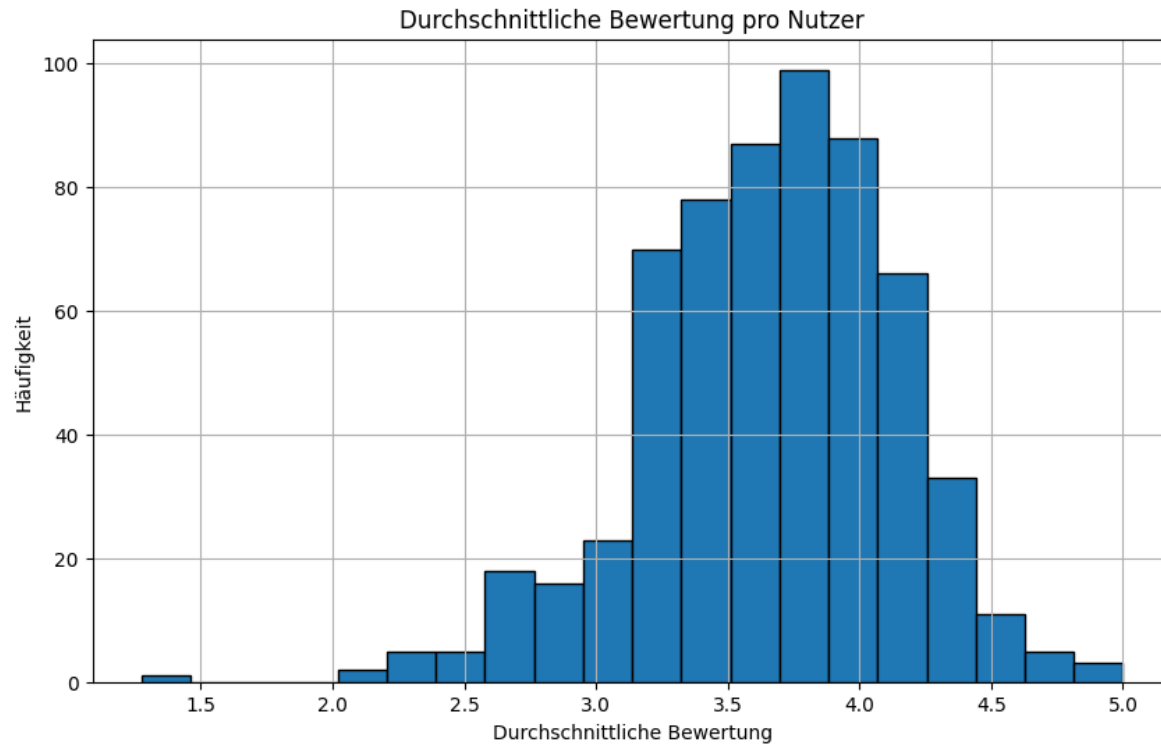
Data
Preparation

Modelling

Evaluation

Deployment

Data Understanding



Business
Understanding

Data
Understanding

Data
Preparation

Modelling

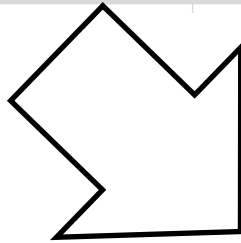
Evaluation

Deployment

Data Preparation

movies.csv

	A	B	C
1	movieId	title	genres
2	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy



	A	B	C	D	E	F	G	H	I	J
1	movieId	title	year	Action	Adventure	Animation	Children	Comedy	Fantasy	Romance
2	1	Toy Story	1995	0	1	1	1	1	1	0

Business
Understanding

Data
Understanding

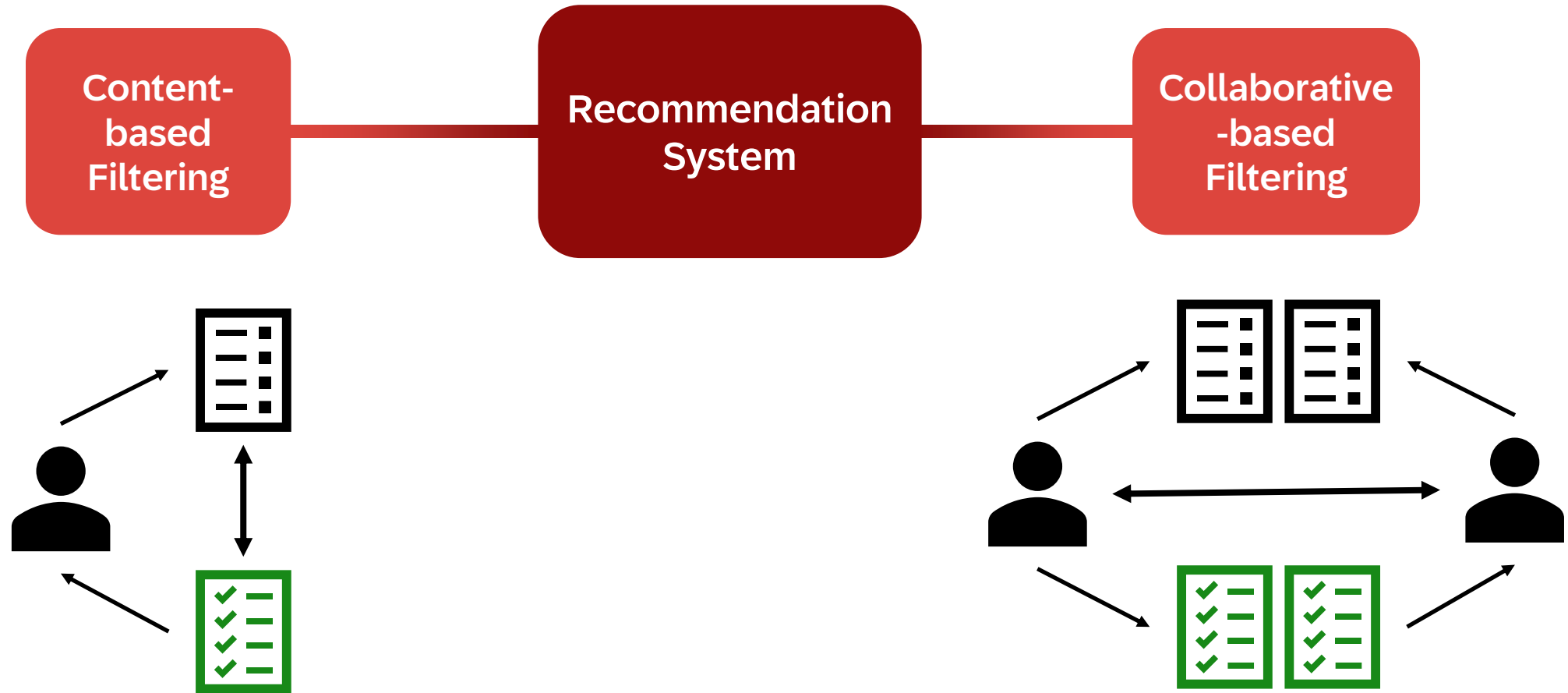
Data
Preparation

Modelling

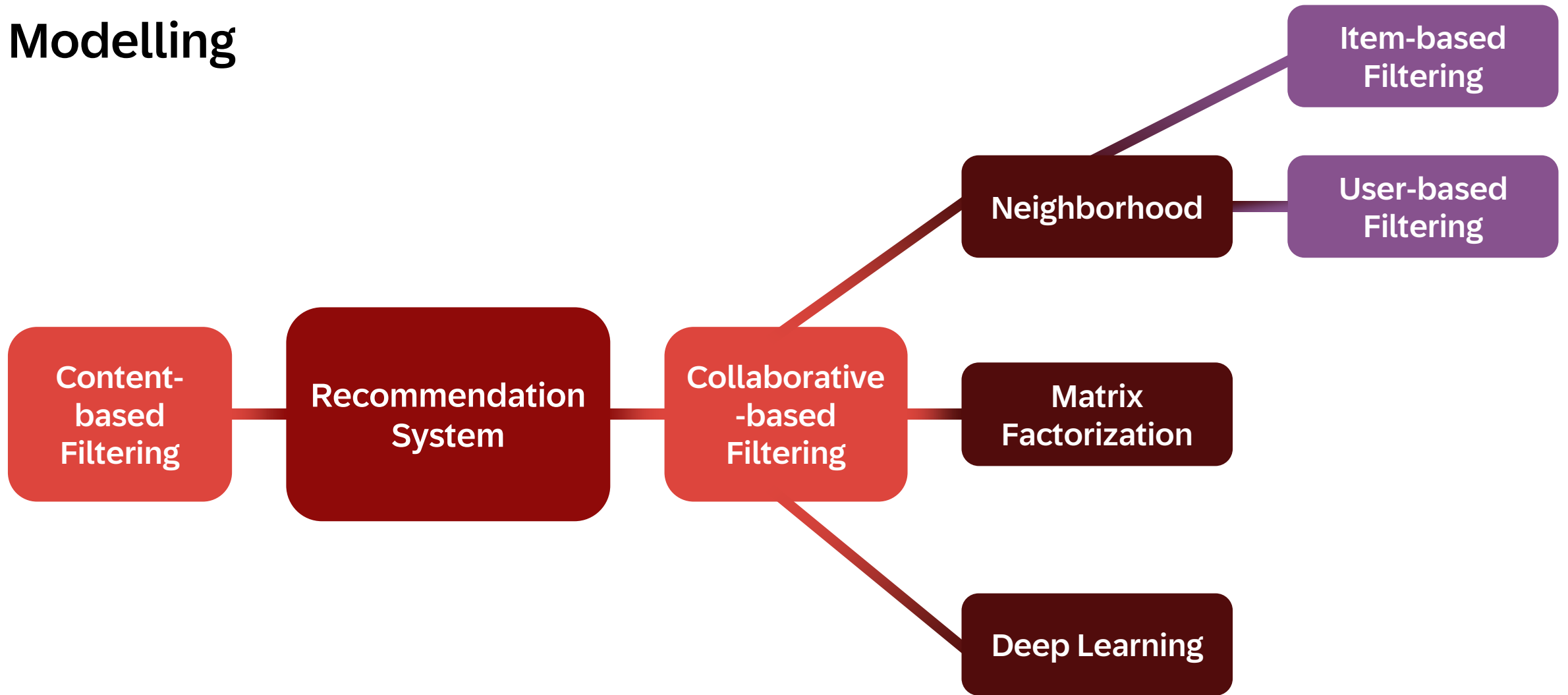
Evaluation

Deployment

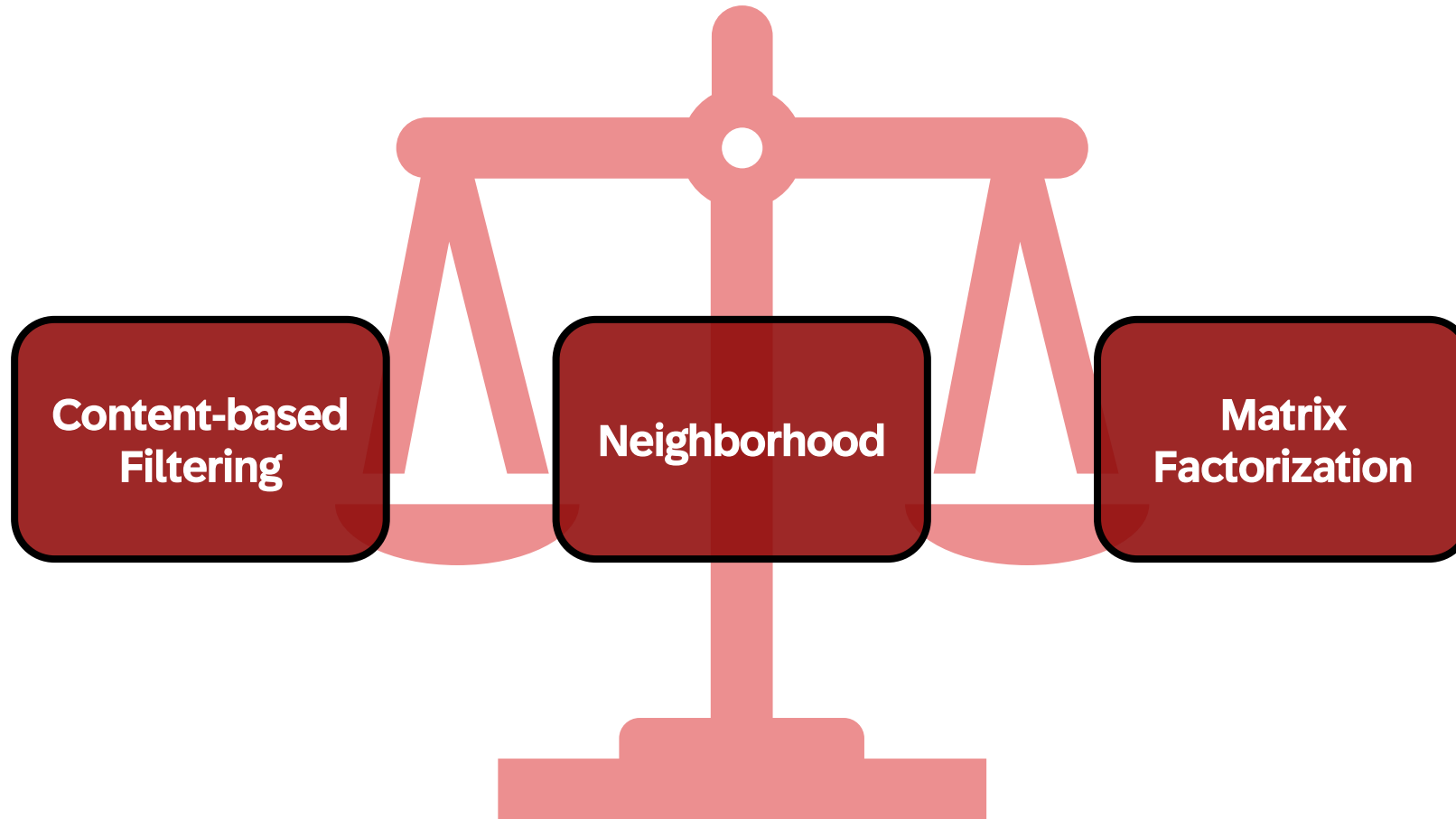
Modelling



Modelling

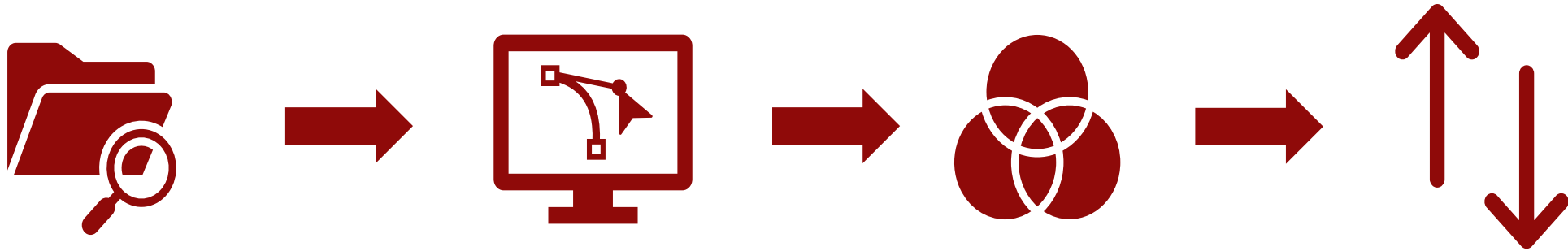


Modelling



Vorgehen

Content-based Filtering



**Merkmalsextraktion
bewerteter Filme**

**Feature-
Vektorisierung**

**Ähnlichkeits-
berechnung**

**Sortierung und
Ausgabe**

Business
Understanding

Data
Understanding

Data
Preparation

Modelling

Evaluation

Deployment

Schritt 1: Merkmalsextraktion

Content-based Filtering

- Filme mit positiver Bewertung (≥ 2.5 Sterne) auswählen
- Falls keine positiven Bewertungen
Auswahl der Filme < 2.5 Sterne
- Genres der Filme aus movies.csv suchen



**Merkmalsextraktion
bewerteter Filme**

Business
Understanding

Data
Understanding

Data
Preparation

Modelling

Evaluation

Deployment

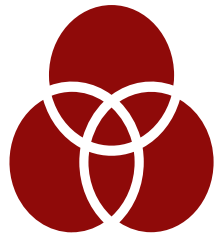
Schritt 2 + 3: Vektorisierung und Ähnlichkeit

Content-based Filtering

Vektorisierung mittels Mittelwerts der Merkmale aller bewerteten Filme = Nutzerprofil



**Feature-
Vektorisierung**



**Ähnlichkeits-
berechnung**

Ähnlichkeitsberechnung zwischen Nutzerprofil
Kinofilm-Merkmalen mithilfe Kosinus-Ähnlichkeit

Business
Understanding

Data
Understanding

Data
Preparation

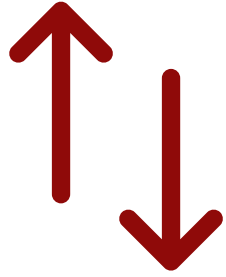
Modelling

Evaluation

Deployment

Schritt 4: Ausgabe

Content-based Filtering



**Sortierung und
Ausgabe**

Sortieren der Filme basierend auf
Ähnlichkeitswerten + Rückgabe mit Scores



Kann auch mit unbekannten Filmen
umgehen!

Business
Understanding

Data
Understanding

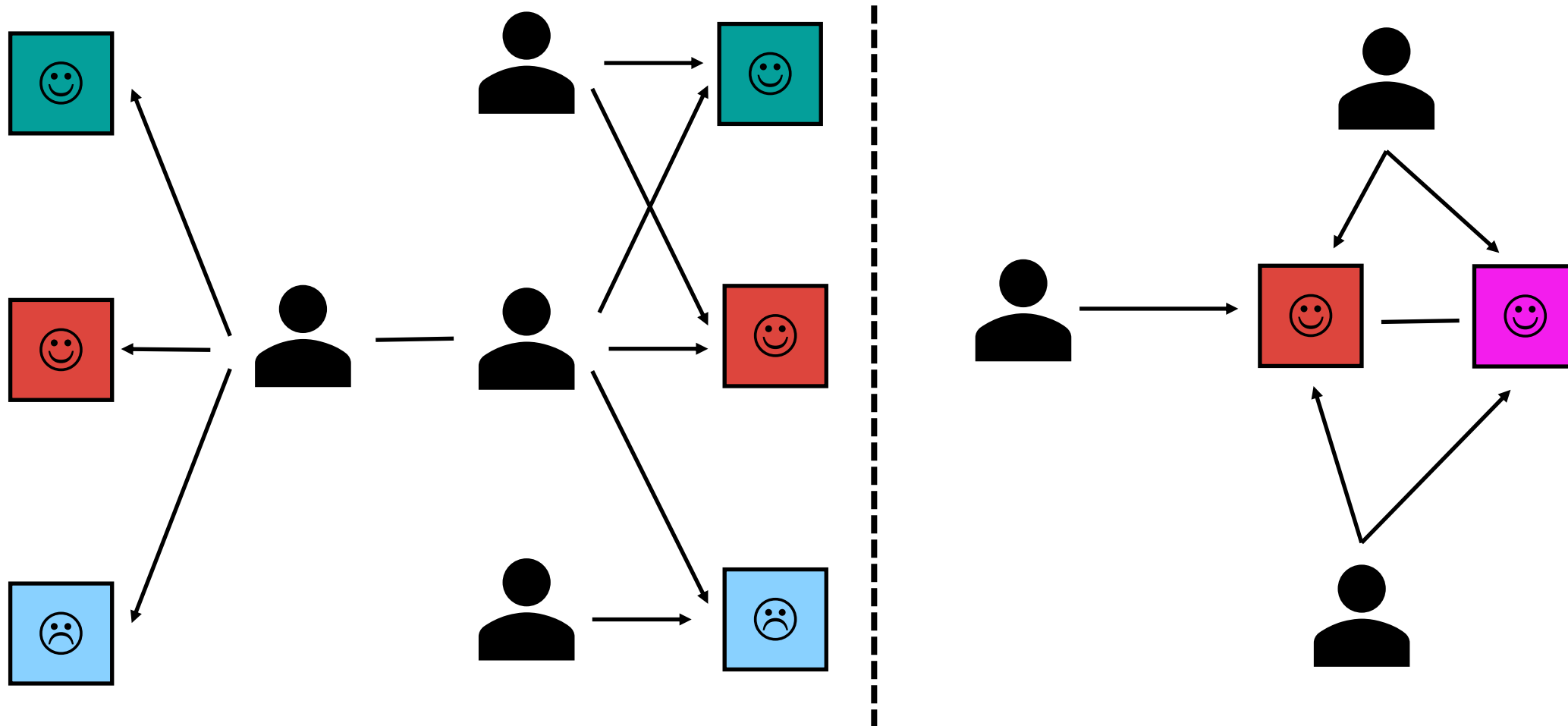
Data
Preparation

Modelling

Evaluation

Deployment

Neighborhood Verfahren - Idee



Business
Understanding

Data
Understanding

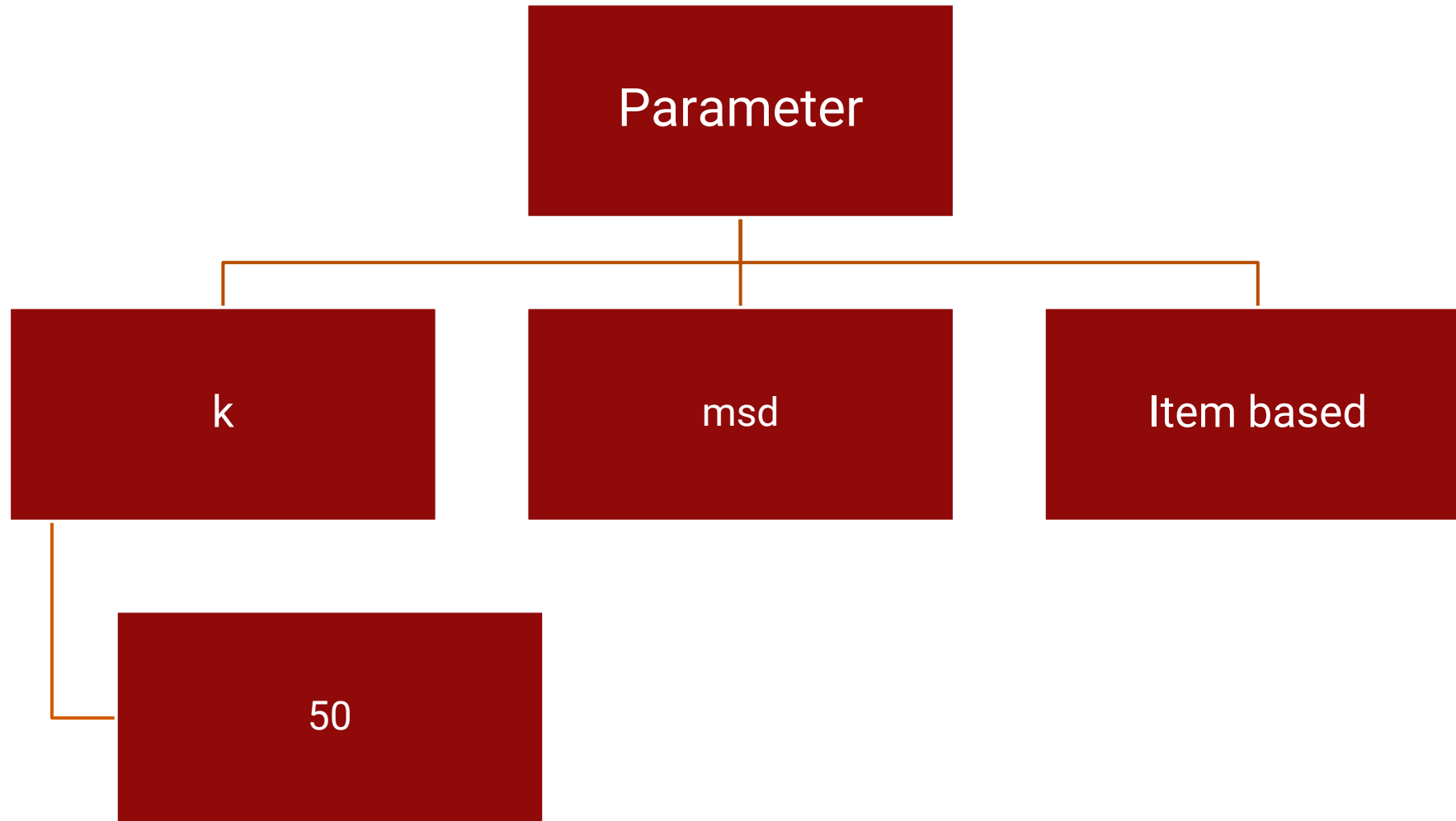
Data
Preparation

Modelling

Evaluation

Deployment

Neighborhood Verfahren – Hyperparameter Tuning



Neighborhood Verfahren – Ähnlichen Nutzer finden

Erstellung einer Ähnlichkeitsmatrix

- Pivot-Tabelle erstellen mit userId als index, movield in der Spalte und ratings in den Zellen

Hinzufügen der neuen Ratings

- Der Neue Nutzer wird zu Pivot-Tabelle hinzugefügt

Berechnung der Ähnlichkeit zu anderen Nutzern:

- Bestimmung der Ähnlichkeit des Nutzers zu allen anderen Nutzern mithilfe der Cosinus Ähnlichkeit

Rückgabe des ähnlichen Nutzers:

- Rückgabe der IDs der ähnlichsten Nutzer für potenzielle Empfehlungen basierend auf deren Vorlieben.

Business
Understanding

Data
Understanding

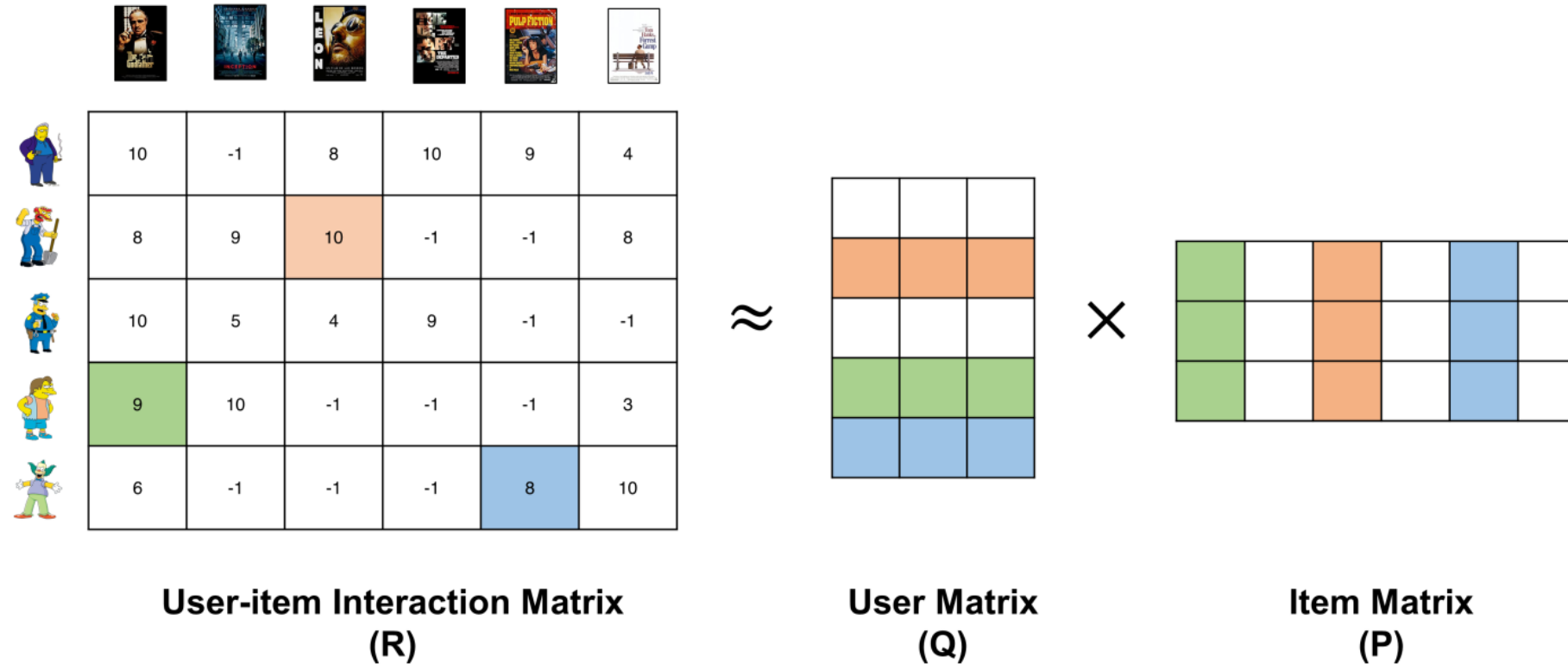
Data
Preparation

Modelling

Evaluation

Deployment

Matrix-Faktorisierung - Einführung



Business
Understanding

Data
Understanding

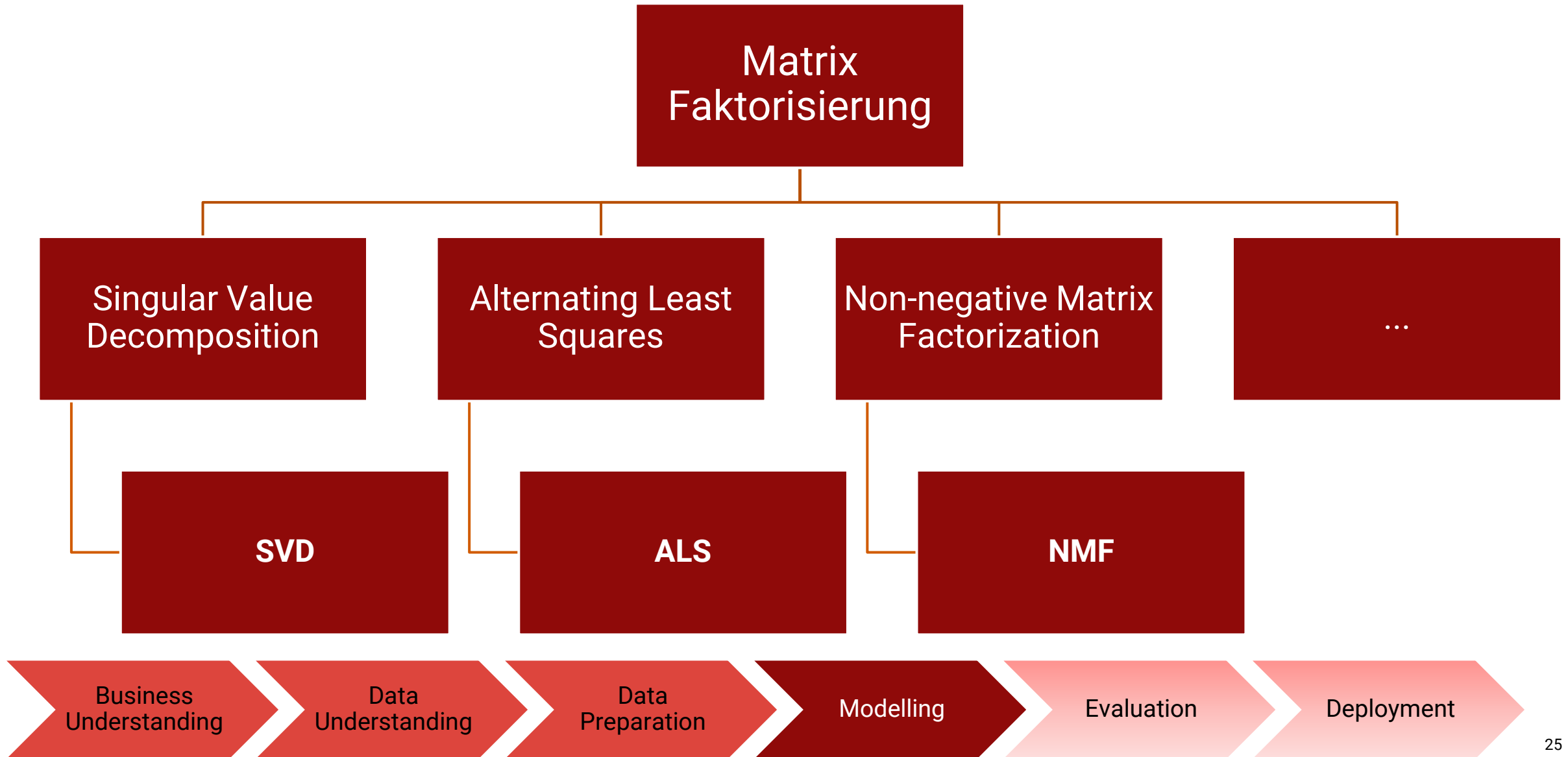
Data
Preparation

Modelling

Evaluation

Deployment

Matrix-Faktorisierung - Methoden



Matrix-Faktorisierung – Bewertungsmetriken

MAE (Mean Absolute Error)

- $MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$

RMSE (Root Mean Square Error)

- $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$

Business
Understanding

Data
Understanding

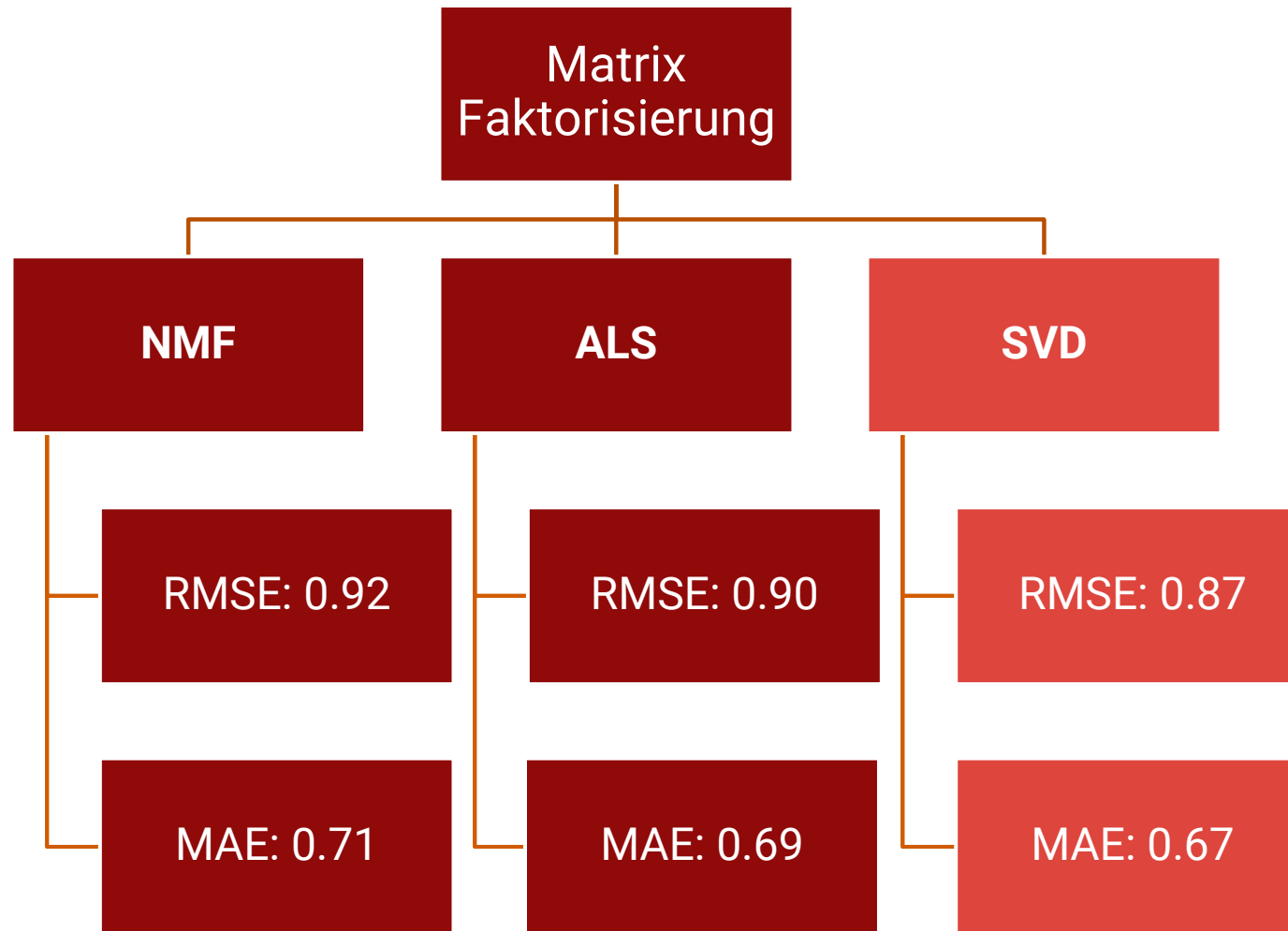
Data
Preparation

Modelling

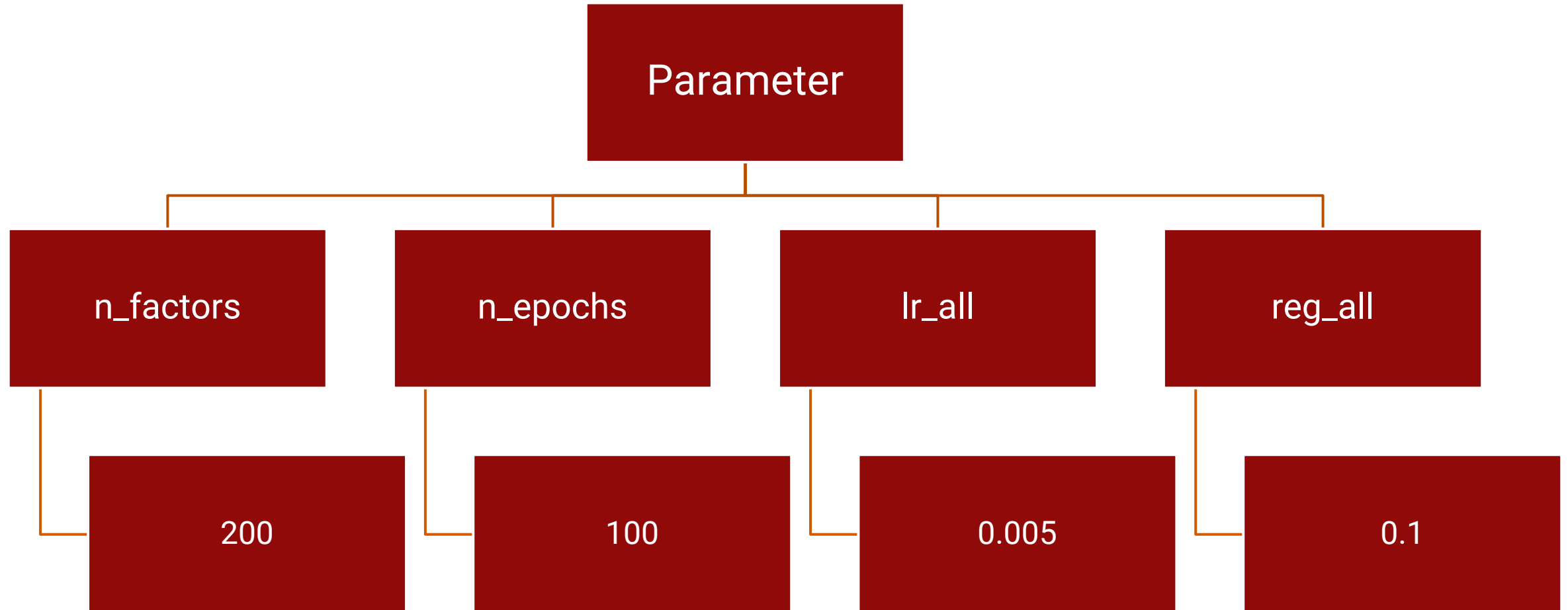
Evaluation

Deployment

Matrix-Faktorisierung - Auswahl



Matrix-Faktorisierung – Hyperparameter Tuning



Matrix-Faktorisierung – Ähnlichen Nutzer finden

Extraktion der latenten Faktoren:

- Gewinnung der latenten Faktoren der bewerteten Filme aus dem trainierten Modell.

Berechnung der Nutzerfaktoren:

- Ableitung der Nutzerfaktoren basierend auf den extrahierten latenten Faktoren und den Nutzerbewertungen.

Berechnung der Ähnlichkeit zu anderen Nutzern:

- Bestimmung der Ähnlichkeit des Nutzers zu allen anderen Nutzern durch das Skalarprodukt der Nutzerfaktoren.

Sortierung der Ähnlichkeiten:

- Sortierung der berechneten Ähnlichkeiten in absteigender Reihenfolge.

Rückgabe des ähnlichen Nutzers:

- Rückgabe der IDs der ähnlichsten Nutzer für potenzielle Empfehlungen basierend auf deren Vorlieben.

Business
Understanding

Data
Understanding

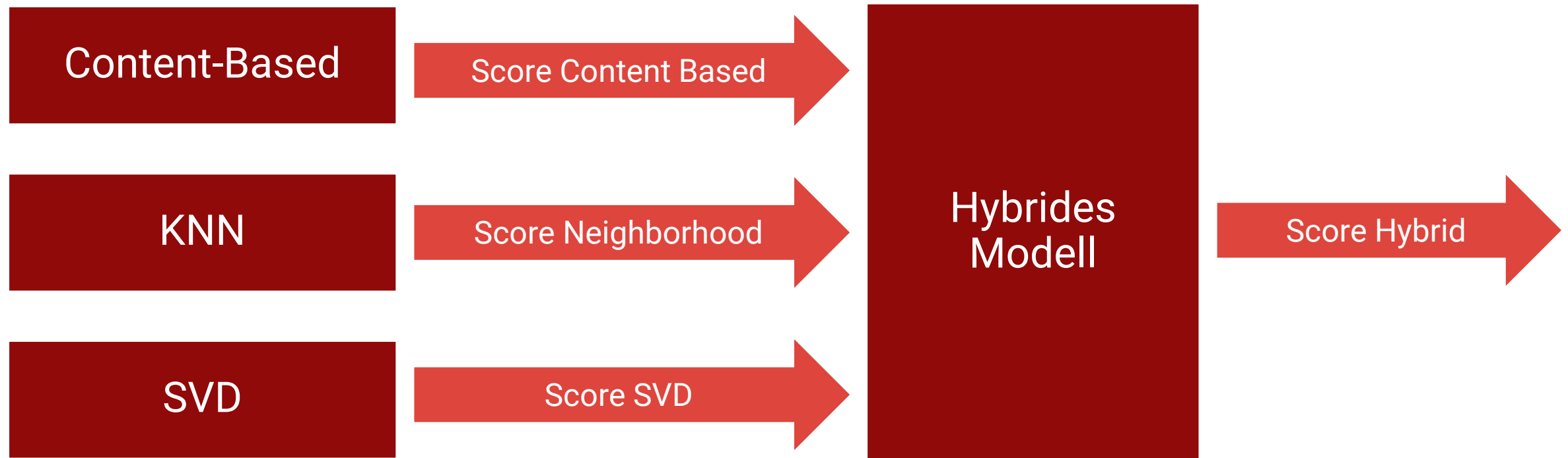
Data
Preparation

Modelling

Evaluation

Deployment

Kombination: Hybrides Modell



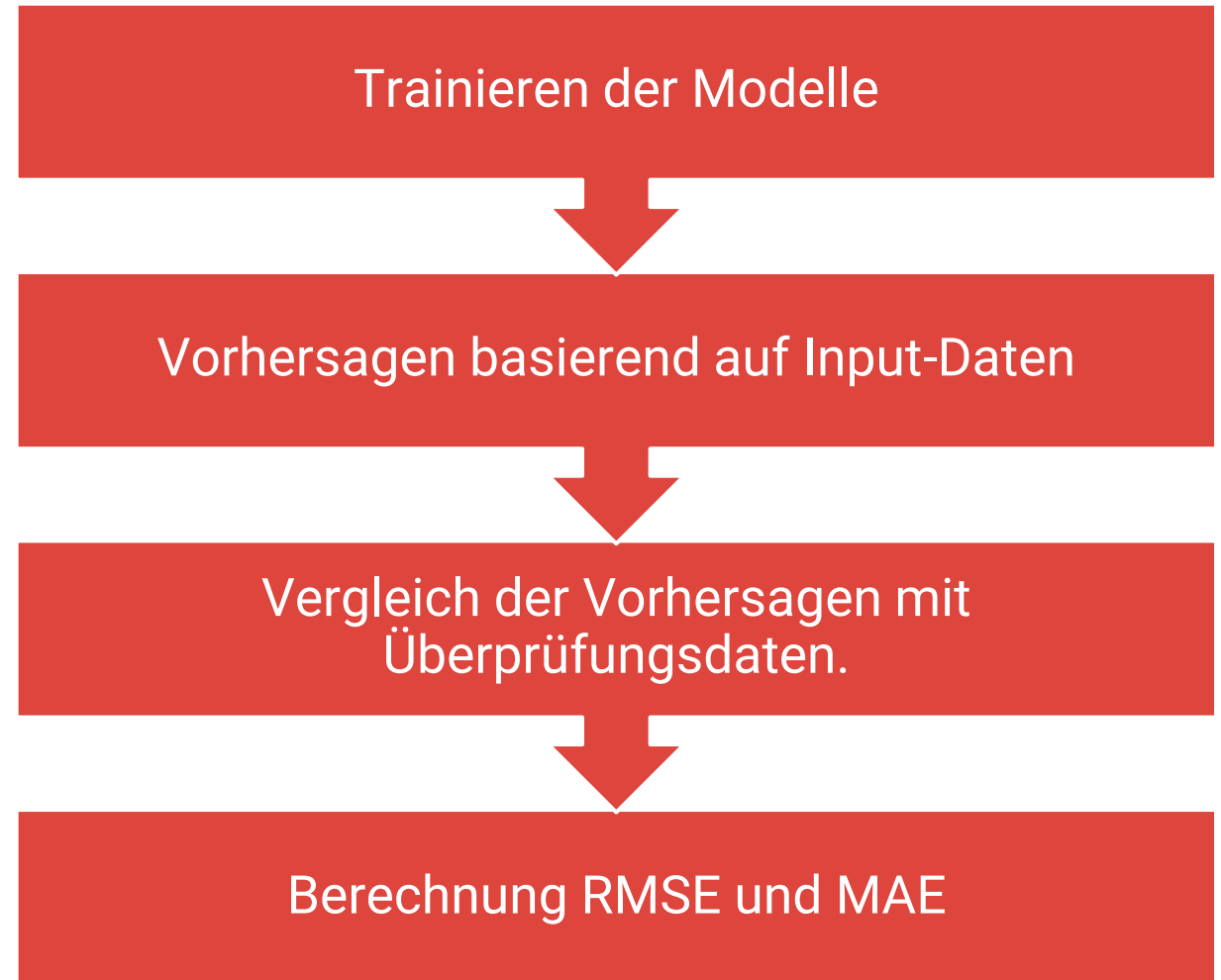
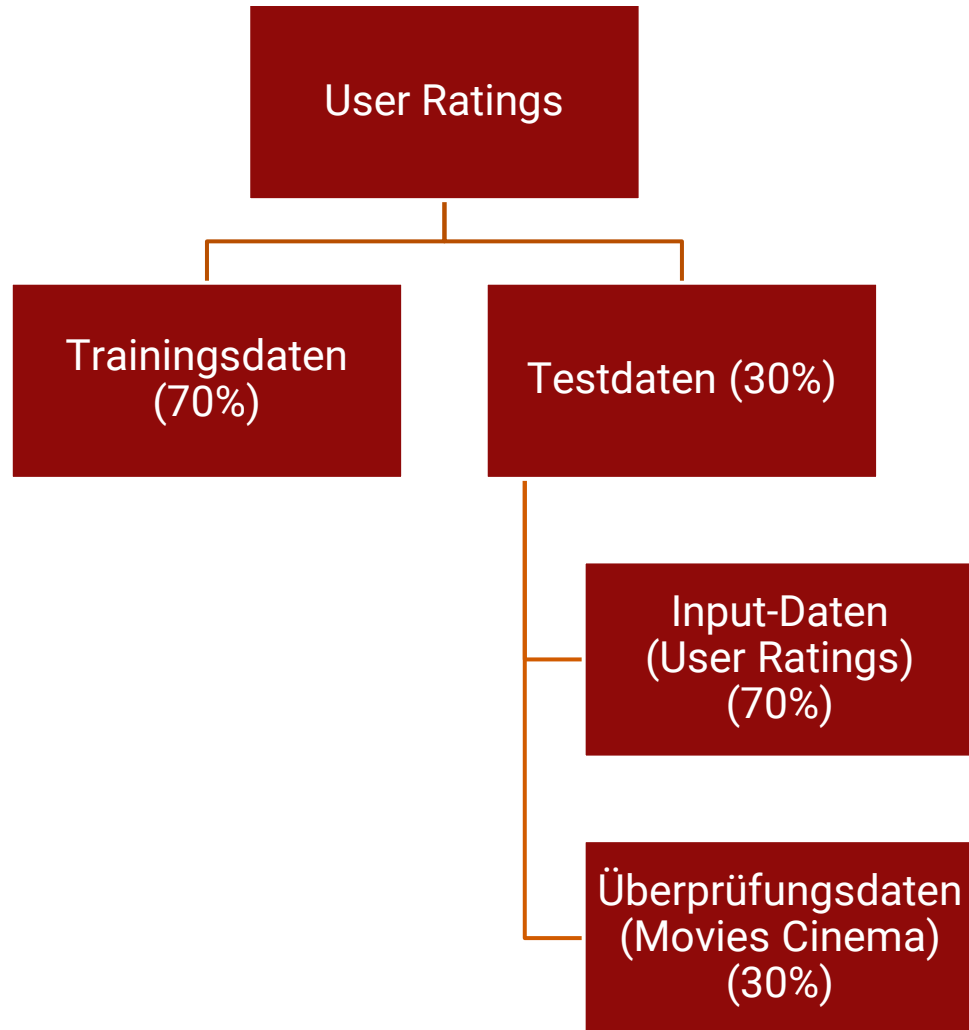
Kombination: Hybrides Modell - Gewichtung

$$\text{Score Hybrid nach Durchschnitt} = \frac{(\text{Score Content Based} + \text{Score Neighborhood} + \text{Score SVD})}{3}$$

$$\text{Score Hybrid mit Gewichtung} = \frac{(\text{Score Content Based} * w1 + \text{Score Neighborhood} * w2 + \text{Score SVD} * w3)}{(w1 + w2 + w3)}$$



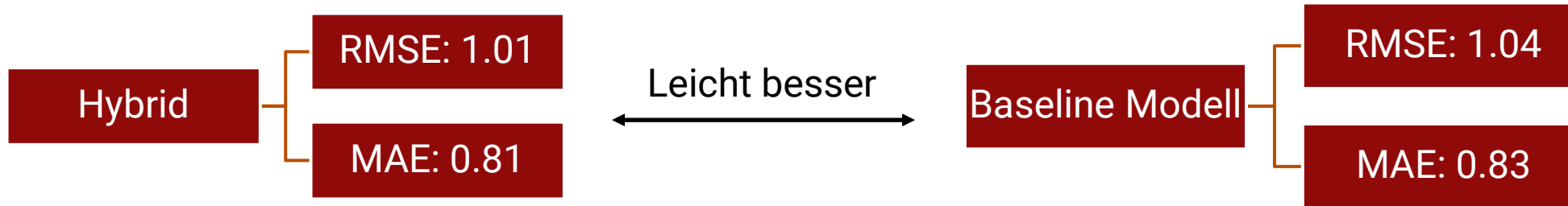
Kombination: Hybrides Modell – Bewertung Vorgehensweise



Kombination: Hybrides Modell – Optimierung der Gewichtung



Evaluation – Bewertung des hybriden Modells



- Mittelwert statisch optimal
- Mittelwert in Realität unaussagekräftig



Geschäftsziele

- Nutzerbindung vertiefen
- Bereitstellung personalisierter Inhalte
- Kompetitive Differentiation

Business
Understanding

Data
Understanding

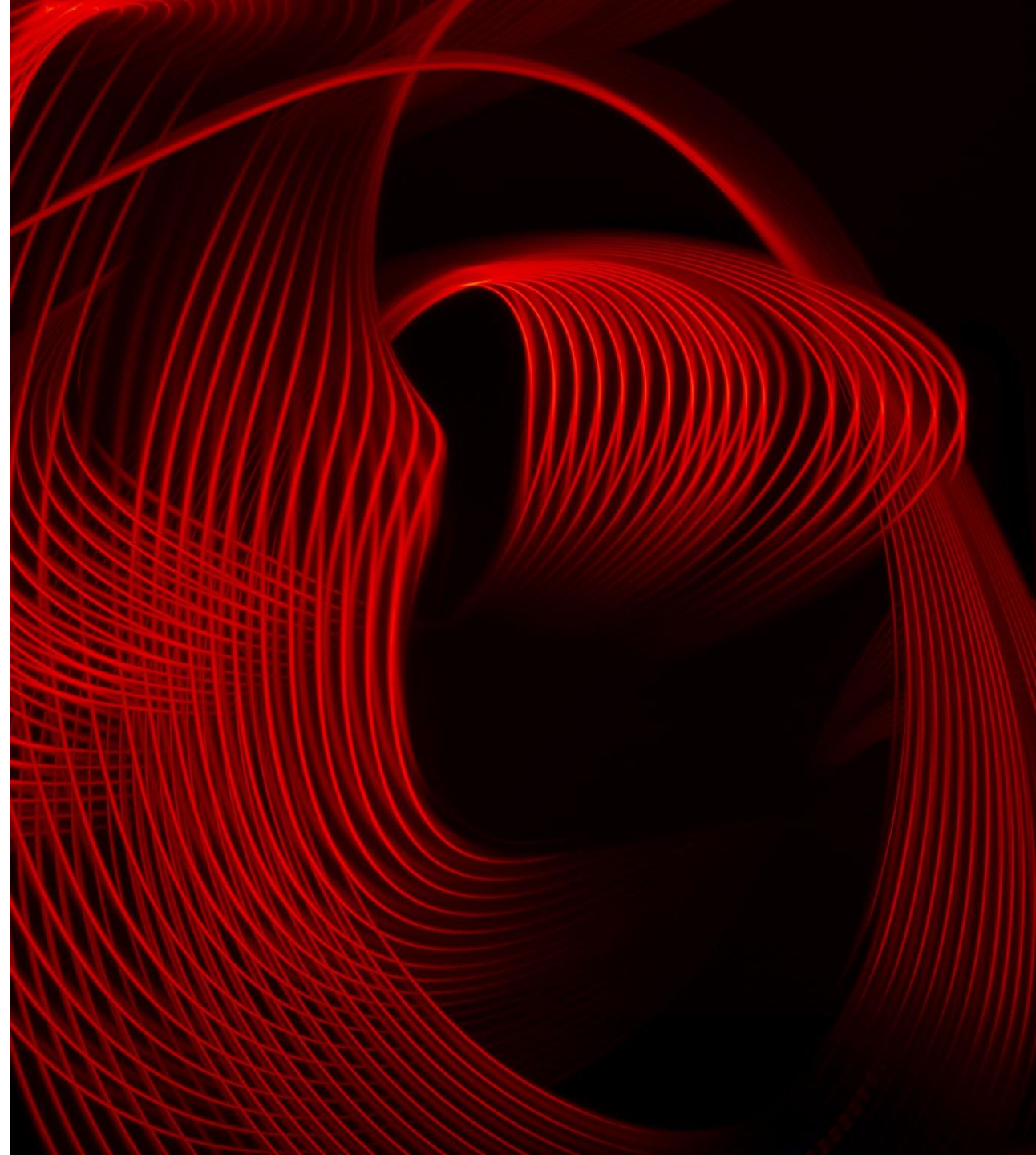
Data
Preparation

Modelling

Evaluation

Deployment

Deployment & Demo



Kritische Würdigung

Nutzerintegration

- Ähnliche Nutzer als Annäherung
- Genauigkeitsverluste (RMSE, MAE)

Eingeschränkte Vorhersagen

- Nur im Modell enthaltene Filme
- Keine neuen Filme (KNN, SVD)
- Content-based Potential (Genres, etc.)

Cold-Start Problem

- SVD und KNN Einschränkungen

Bias in den Daten

- Nette und kritische Nutzende

Skalierbarkeit

- Problem bei großen Datenmengen
- Kino idR max. 20 Filme -> kein Problem

Ausblick

- Deep Learning Potential

A perspective view of a movie theater. The foreground is filled with rows of red upholstered seats. In the background, a large white screen is illuminated, displaying the words "The End" in a black, elegant cursive script. The theater walls are dark, and the overall lighting is dim, focusing on the screen.

The End

URLS:

Api Doku:

<https://moviemate.mabu2807.de/>

Kino Webseite:

<https://cinetastisch.mabu2807.de/>