



## Big Data Analytics – Methoden und Anwendungen

Prof. Dr. Mathias Klier

**Lara Frost**  
**Chiara Schwenke**

Institut für Business Analytics

02.06.2025

**Einführung in die Gruppenprojekte**

# Organisatorische Kurzbeschreibung & Ablauf der Gruppenprojekte

**Mo  
02.06.**

- Ausgabe der Aufgabenstellung und der Datensätze
- Start der Anmeldung für das Gruppenprojekt (Gruppenbildung ab **12:00** Uhr)
  - siehe Gruppenvergabetool auf Moodle
  - die maximale Gruppengröße beträgt vier Personen

**Mo  
09.06.**

- Um **23:59** Uhr Anmeldeschluss für die Anmeldung zum Gruppenprojekt, ab dann sind die Gruppen fix und können nicht mehr geändert/getauscht werden!

**Do  
17.07.**

- Ende der Bearbeitungszeit
- Abgabe des .ipynb Files sowie der Präsentationsfolien per E-Mail an: [lara.frost@uni-ulm.de](mailto:lara.frost@uni-ulm.de) bis **e.o.d.**

**Mo  
21.07.**

- Präsentation der Gruppenprojekte
- Nennung der Gewinner und Bekanntgabe der erreichten Bonuspunkte 😊

# Inhaltliche Kurzbeschreibung der Gruppenprojekte

## Ausgangssituation

In digitalen Plattformen mit umfangreichen Produkt-, Inhalts- oder Angebotsportfolios fällt es Nutzer:innen zunehmend schwer, relevante Optionen zu finden. Die Folgen:

- sinkende Interaktion der Nutzer:innen.
- geringere Kundenzufriedenheit.
- abnehmende Kundenbindung.

## Herausforderungen

- Teils sehr geringe Nutzeraktivität.
- Oft werden nur Bestseller gefunden, keine persönlichen Empfehlungen.

## Zielsetzung

- Entwicklung eines Recommender Systems in Python mit Fokus auf Nutzerzufriedenheit, Bewertungsvorhersage und systematischer Evaluierung.
- Hierzu können bspw. die Bausteine der Methoden des Collaborative Filterings sowie des Content-Based Filterings angepasst und kombiniert werden, um die Empfehlungsqualität zu verbessern.
- Vorbereitung einer überzeugenden Präsentation für die jeweiligen Vorstände.

# Aufgabenstellung für die Gruppenprojekte

- Entwickeln Sie auf Basis Ihrer Kenntnisse zu Collaborative und Content-Based Filtering das **bestmögliche Recommender System in Python**. Entwickeln und optimieren Sie Ihr Recommender System, indem Sie bisher vorgestellte Methoden **ergänzen, kombinieren, weiterentwickeln, anpassen, konfigurieren, ...**
- Ihre **Herausforderung**:
  - 1 Das bestmögliche Recommender System = **höchste Empfehlungsgüte** (gemessen am mittleren absoluten Fehler in Sternen, MAE) unter allen Gruppen Ihres Datensatzes.
  - 2 Eine **überzeugende Präsentation**, in der Sie Ihre **Idee** und Ihre **Ergebnisse** vorstellen und den **Vorstand** von Ihrem **Recommender System** und Ihrer **Arbeit** überzeugen!  
→ Ihrer **Kreativität** sind hier keine Grenzen gesetzt!
- **Abgabe bis spätestens 17.07.** an [lara.frost@uni-ulm.de](mailto:lara.frost@uni-ulm.de): .ipynb file & Präsentation in MS PowerPoint (**5 - 10 Folien**) mit der Sie Ihren Ansatz Methode(n), Begründung, Ergebnisse / Empfehlungsgüte in der Veranstaltung in max. **8 - 10** Minuten präsentieren können.

## Szenario 1 – FlixNet

- Mit dem Ziel, die Nutzerzufriedenheit auf einer Streaming-Plattform zu verbessern, stellt **FlixNet** Ihnen Nutzungs- und Bewertungsdaten zu seinem Film- und Serienkatalog bereit.
- Der Datensatz enthält Informationen zu Filmen, die mindestens **5** Bewertungen erhalten haben. Zudem sind die Bewertungen aktiver Nutzer:innen enthalten, die **19** oder mehr Filme bewertet haben.

### Informationen zum Datensatz



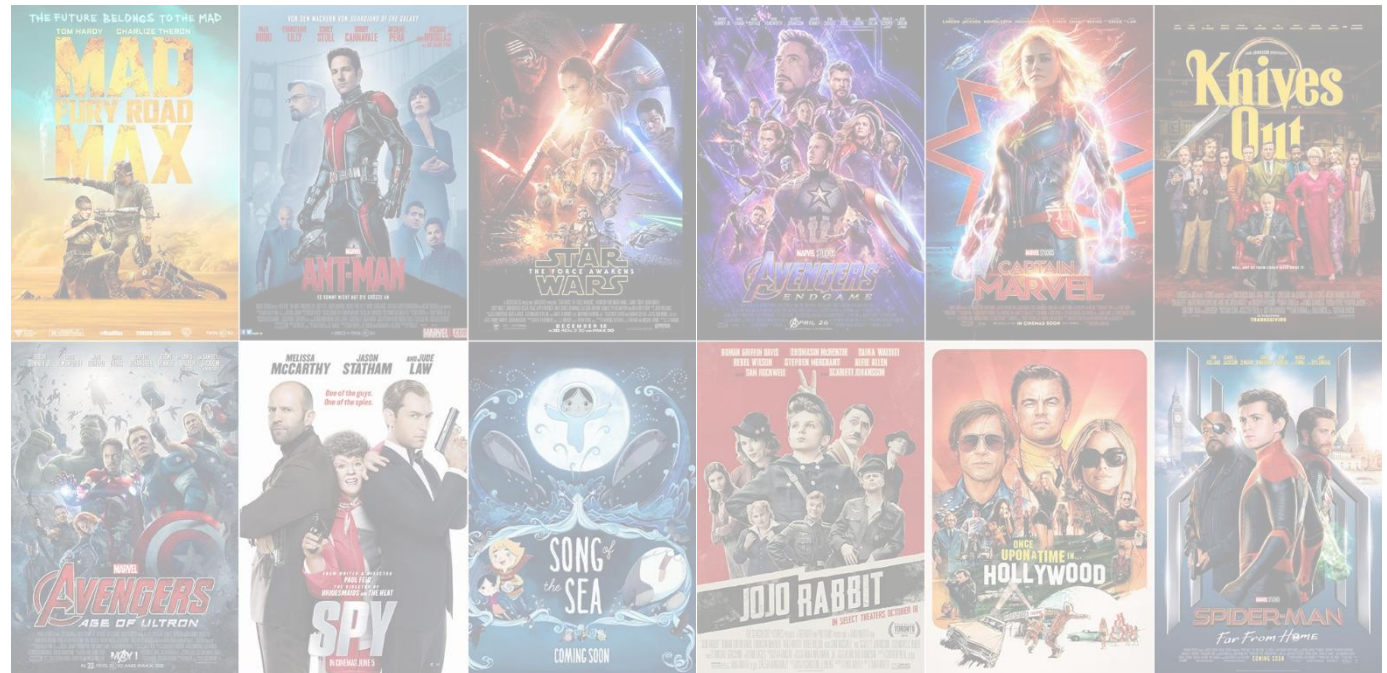
**User:** 800



**Items:** 788



**Skala:** 0 – 5 Sterne





## Szenario 2 – SephoBay

- Angetrieben vom Wunsch, die Kundenbindung im Onlinehandel zu steigern, stellt Ihnen **SephoBay** anonymisierte Bewertungsdaten zu seinen Produkten zur Verfügung.
- Enthalten sind Informationen zu Drogerie-Produkten aus diversen Kategorien, die von mindestens **5** Käufer:innen bewertet wurden sowie die Bewertungen von Nutzer:innen, die **17** oder mehr Produkte bewertet haben.

### Informationen zum Datensatz



**User:** 798



**Items:** 622



**Skala:** 0 – 5 Sterne



## Szenario 3 – Bibliona

- Inspiriert durch die geplante Einführung einer personalisierten Empfehlungsfunktion stellt Ihnen **Bibliona** einen Datensatz mit Leserbewertungen zur Verfügung.
- Der Datensatz umfasst Informationen zu Büchern, die mindestens **9** Mal bewertet wurden, sowie anonymisierte Bewertungen der Nutzer:innen, die mindestens **5** Bücher bewertet haben.

### Informationen zum Datensatz



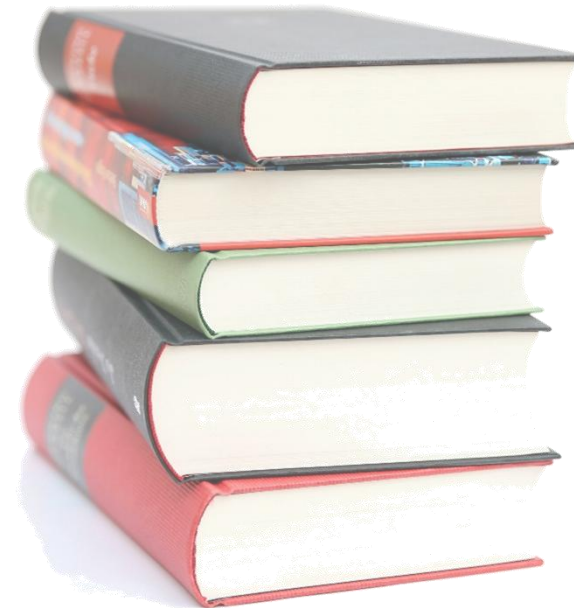
**User: 798**



**Items: 781**



**Skala: 0 – 10 Sterne**



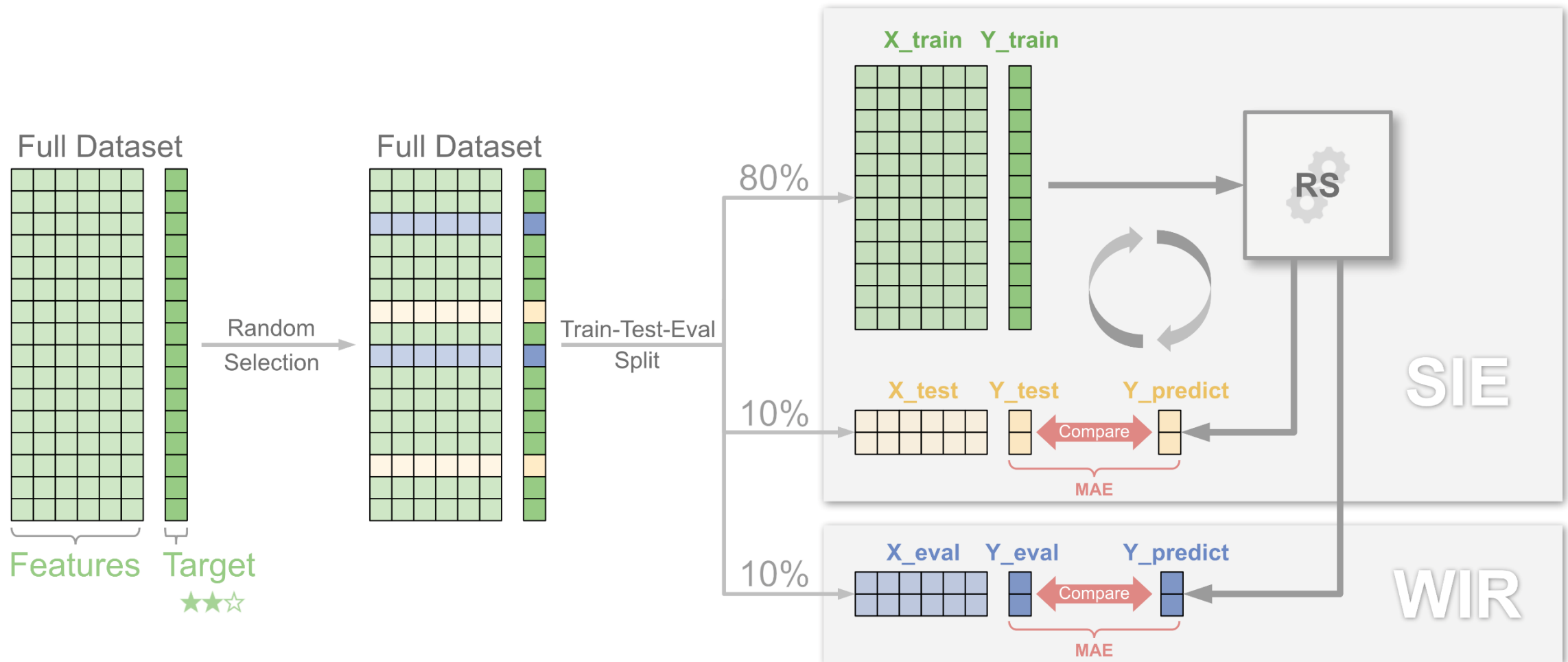
## Datensätze und Testdaten

- Zu jedem Szenario sind in Moodle die Bewertungen der betrachteten Nutzer:innen und Unternehmen in den Dateien **Bewertungsmatrix.csv** und **Ratings.csv** hinterlegt, sowie in **Itemprofile.csv** die Itemprofile aller im jeweiligen Szenario betrachteten Unternehmen.
- Diese Dateien haben dieselbe Struktur wie die gleichnamigen Tabellen aus Übung 3 und 4.
- Zudem ist zu jedem Szenario ein **Testdatensatz.csv** hinterlegt. Hierin sind jeweils ca. 10% aller Bewertungen enthalten, die aus dem Datensatz entfernt wurden.
- Durch die Vorhersage der Testdaten können Sie die **Güte** ihres Ansatzes messen!
- Weitere ca. 10% aller Bewertungen wurden von uns entfernt und dienen als weiterer Testdatensatz für die **Evaluation**.



## Evaluation: Messung der Güte durch die Vorhersage von Testdaten (I)

- Als Ihre „Auftraggeber“ evaluieren wir die Empfehlungsgüte Ihres Recommender Systems anhand streng unter Verschluss gehaltener **Evaluationsdaten**.



## Evaluation: Messung der Güte durch die Vorhersage von Testdaten (II)

### Problem

- Vorhersage muss für jeden einzelnen Testdateneintrag (in Sternen) bestimmt werden
- Manuell sehr aufwändig, da die Eingabeparameter der des Recommender Systems (z.B. User und Item) immer neu gesetzt werden müssen

### Lösung

- Erzeuge eine Schleife, die automatisiert das Recommender System für alle Testdaten aufruft, die Eingabeparameter entsprechend ändert und die Ergebnisse sammelt.

### Ihr Recommender System

- Collaborative Filtering
- Content-Based Filtering
- Kombinationen
- Erweiterungen
- Eigener Ansatz
- ...



**Wie gut schneiden Sie im Vergleich zu den konkurrierenden Gruppen ab?**

## Präsentation für den Vorstand

- Bereiten Sie eine Präsentation in MS PowerPoint vor, die ca. **5 -10 Folien** umfasst und die Sie am **17.07.** in **8-10 Minuten** präsentieren können.
  - Veranschaulichen und visualisieren Sie Ihre **Idee**, sodass diese **leicht verständlich** ist.
  - Bereiten Sie Ihre **Ergebnisse** auf, die Sie mit Ihrem Recommender System auf dem **Testdatensatz** erzielt haben.



**Hat Ihre Präsentation das Potenzial, den Vorstand zu überzeugen und Ihnen eine Beförderung einzubringen?**

## Jetzt sind Sie an der Reihe 😊

- Entwickeln Sie ein **Recommender System**, das die Zielgröße **mittlerer absoluter Fehler (MAE)** für die Testdaten minimiert. Hierfür muss ihr Recommender System eine möglichst exakte **Vorhersage** in „**Sternen**“ bieten.\*
- Seien Sie **kreativ**, **kombinieren** Sie bestehende Ansätze, **erweitern** Sie sie und entdecken Sie **neue Funktionen** in Python.
- **Überzeugen** Sie mit Ihrer **Präsentation**.
- Nutzen Sie bei **Fragen** das **Moodle-Nachrichtenforum**.

# Viel Erfolg!

\*Hinweis: Content-Based Filtering kann also nicht direkt angewendet werden.

Finden Sie z.B. eine kreative Möglichkeit, die Ergebnisse trotzdem mit einfließen zu lassen?