**Ziel von Sprint 2: Datenaufbereitung**

**Inhalte des Sprint Planung 2: 07.07-20.07**

1. **Sprint-Ziel festlegen (gemeinsam im Team):**

* Große CSV-Datei endgültig bereinigen/vorbereiten.
* Autorrating systematisieren
* Genre systematisieren

Fragen:

Autor\_rating fehlt oder nicht richtig, erl

Book\_avarage\_rating fehlt oder nicht richtig, erl

Book\_rating ount fehlt oder nicht richtig, erl.

Ausgabeverlag-Name systematisieren

* Daten bereinigen
* Daten analysieren
* Kategorische Werten in numerische umwandeln
* Fertige csv Datei mit 2 verschiedenen Modellen trainieren: 1.RandomForestClassifier(random\_state=42, class\_weight={0:1, 1:2} und mit GridSearchCV

2.LogisticRegression(max\_iter=1000, random\_state=42)

auswerten mit Confusion smatrix

1. **Gemeinsame streamlit\_app.py fertigstellen**

* Nur grob als platzhalter

1. **Teils-Themen ausarbeiten:**

* Voraussage mit ML , ob eine neues Buch verfilmt wird(LW)
* Buchempfehlung(AW)
* Kommerzielle Marktanalyse (JS)

**2. Aufgaben aus dem Product Backlog auswählen (und aufteilen)**

| **Aufgabe (User Story / Task)** | **Zuständig** | **Aufwand** |
| --- | --- | --- |
| CSV Datei vorbereiten | Person | LW |
| Empfehlung- Daten | Person | AW |
| Kommerzielle Daten | Person | JS |
|  |  |  |