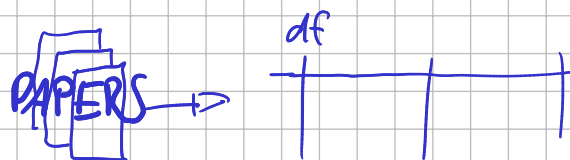


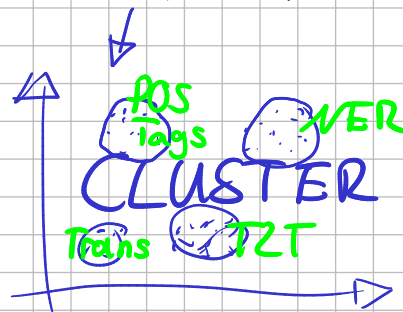
Zumindest Vorgehensweise insgesamt

1 Cluster bilden (Sentence BERT)



Annahme: Cluster entsprechen Themen / Themengebieten

⇒ Cluster

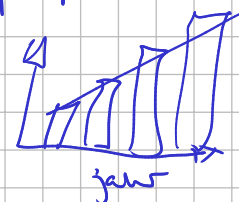
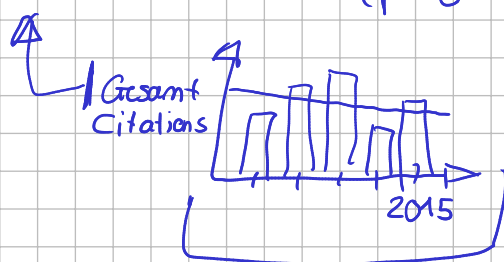


2 Cluster analysieren (komplett historische Daten)

mehrere Papers / 1 cluster

Summe Citation Counts (pro Jahr)

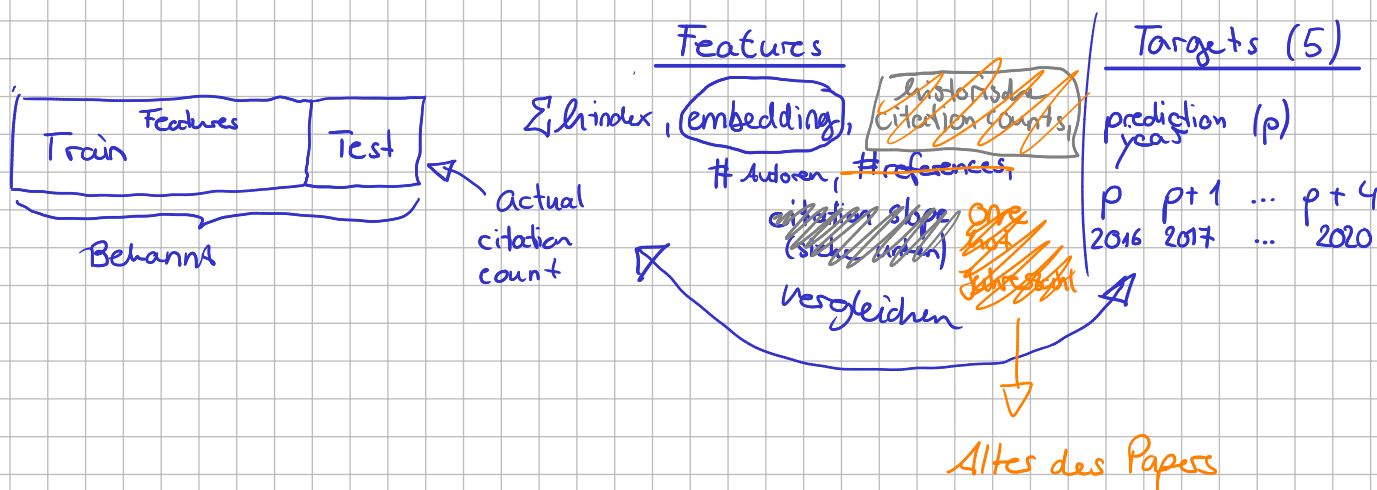
Anzahl Papers pro Jahr



Statistik → Trends
Citation count Entwicklung

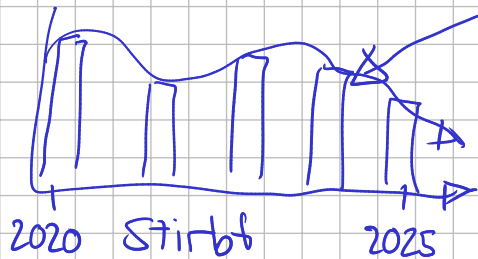
3 Zukünftige Citations vorhersagen

- Model erstellen → Architektur:
- Model trainieren & evaluieren



4 Clusterweise Trends visualisieren

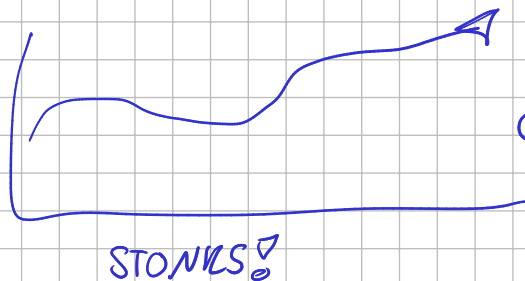
C0



Summe Citation Counts der Paper in Cluster C0 im Jahr 2024

Σ (Output vom Model)

C1



C1 $\begin{bmatrix} - \\ - \\ - \end{bmatrix}$
C2 $\begin{bmatrix} - \\ - \\ - \end{bmatrix}$

\rightarrow $\begin{bmatrix} ? \end{bmatrix}$ \rightarrow

Citation 2025

$\begin{bmatrix} - \\ - \\ - \end{bmatrix} \begin{bmatrix} - \\ - \\ - \end{bmatrix} \begin{bmatrix} - \\ - \\ - \end{bmatrix} \begin{bmatrix} - \\ - \\ - \end{bmatrix}$
2024
 \downarrow
 Σ

Conclusion oder Wie wir es machen wollen :

Inputs (Features):

- Alter des Papers (in Jahren)
- Embedding (Abstract, Volltext?)
- ^{aktueller} h-index der Autoren (irgendwie zusammengefasst)
- #Autoren

Vorteil: • Paper kann mehrfach zum Trainieren benutzt werden:

Bsp Publikation: 1980

~ 1980 - 1984, 1985 - 1989, ..., 2016 - 2020

- Auch brandneue Paper können predicted werden

Output (Targets): Citation counts für die nächsten 5 (5+) Jahre

Wenn mehrere Paper aus einem Topic predicted werden, kann damit ein Trend für das gesamte Topic erstellt werden.

Um Trends herauszubekommen, alle Paper mit ihrem aktuellen Alter durch das kleinste Model jagen :)