

Recommendation System

Building an hybrid recommender system for documents with NLP integration

1



First Steps

- Reads PDF files contained in folders named as their corresponding categories within a single folder
- Predict documents categories using a NLP classifier (Doc2Vec, BERT, etc...)

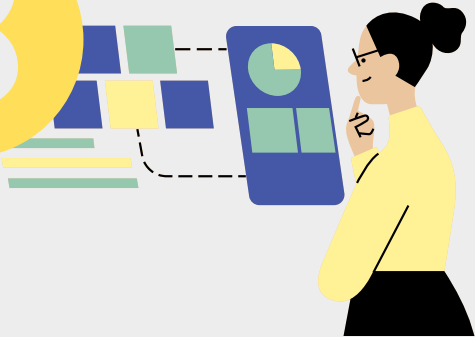
2



Outdoing the cold-start

- Append the category belonging to the last downloaded pdf to user matrix for all users
- Calculate KNN/cosine similarity to find top 10 analogous users
- Define category with most recurrent occurrences
- Recommend 3 pdf of top 10 users that share category

3



Tailoring users' tastes

After user's first pdf download, perform dot product comparison with corresponding embedding matrix and recommend top 3 pdf with highest values.

4



What's next?

- When a user downloads more than a single pdf, perform embedding vector pooling (or calculate mean)
- Perform dot product of mean embedding with embeddings in a matrix containing all categories
- This could tailor recommendations even further to the users temporary preferences

