Project Proposal: AI Book Recommendations

Problem Selection and Rationale:

Choosing a good book from the millions available can feel like an impossible task. Moreover, books stores need algorithms that can seamlessly and accurately match readers with the books they are most willing to purchase. We aim to address this issue through testing Transformers4Rec and SVD models designed to match readers with books. We will do this by prompting readers/users to provide a brief book review (3-5 sentences) of a recently enjoyed book and use this input to generate a short list of books they are likely to enjoy, based on our database of reviews and user profile data.

Dataset:

For this project, we will utilize the Amazon Book Reviews dataset: https://www.kaggle.com/datasets/mohamedbakhet/amazon-books-reviews. This dataset offers almost 3 GB of data on book reviews and book ratings.

NLP Methods:

We will use Transformer4Rec models to generate book recommendations based on user input (in the form of a short book review of a recently enjoyed book). Model output and performance will be compared to a classical SVD model.

Packages:

- NLTK and spacy for text preprocessing, sentence tokenization, and feature extraction.

- sklearn for vectorization, Singular Value Decomposition (SVD), and Latent Semantic Analysis (LSA).

- transformers4rec for transformer model.

NLP Tasks:

1. Explore and preprocess the book review text data, including cleaning and tokenization.

2. Implement SVD models and Transformer4Rec models with book review data as the input and book recommendations as the output.

3. Evaluate the performance of the models and analyze the results.

Performance Evaluation:

Performance of the models described above will be evaluated based on various metrics, including precision, recall, and F1-scores.

Project Schedule:

1. Week 1: Data preprocessing and exploration. Implement initial Transformer4Rec and SVD models.

2. Week 2: Continue developing the Transformer4Rec and SVD models.

3. Week 3: Test model performance and design Streamlit App interface.

4. Week 4: Document the findings and insights in a final project report and prepare the presentation.