|  |  |
| --- | --- |
| **Topic :** | **BOOK RECOMMENDATION SYSTEM** |
| **Document Type:** | **Low-Level-Design (LLD)** |
| **Author :** | **Arpan Kumar Sasmal** |
| **Date: 10/09/2024** | **PW Skills: Full Stack Data Science Pro** |

**1. Introduction:**

**The Book Recommendation System provides personalized book recommendations using Natural Language Processing (NLP) and dimensionality reduction techniques. By analyzing book characteristics such as titles, authors, genres, and physical attributes, the system aims to suggest books that match user preferences and enhance their reading experience.**

**2. Problem Statement:**

**As the number of books grows, users struggle to find books aligned with their tastes. A recommendation system is required to analyze book features and provide tailored suggestions. The goal is to predict which books a user might enjoy based on their preferences and the characteristics of available books.**

**3. Dataset Information**

**This dataset contains information about books with the following variables:**

* **Title: The name of the book.**
* **Author: The author of the book.**
* **Genre: The main genre the book falls into.**
* **Subgenre: The subcategory within the genre.**
* **Publisher: The company or entity that published the book.**
* **Height: The physical height of the book in centimeters.**

**4. Architecture Description:**

**4.1 Data Description:**

**The dataset is available at** [**GitHub Books Dataset**](https://gist.github.com/jaidevd/23aef12e9bf56c618c41)**. This dataset includes columns such as Title, Author, Genre, Subgenre, Publisher, and Height, which provide comprehensive information about each book for recommendation purposes.**

**4.2 Data Preprocessing:**

**Libraries: Import essential libraries such as Pandas, Scikit-learn, Numpy, Matplotlib, Seaborn, and Streamlit. - Loading Data: Read the dataset and handle missing values, impute necessary fields, and clean the data. - Feature Engineering: Rearrange book titles, scale the Height column, and combine features for recommendation.**

**4.3 Data Analysis:**

**EDA: Inspect the dataset, generate summary statistics, check for missing values, and create visualizations for better understanding (e.g., distribution of genres, word cloud for titles).**

**4.4 Data Transformation (Feature Engineering):**

**TF-IDF Vectorization: Convert combined features into numerical format using TF-IDF. - SVD: Apply Singular Value Decomposition (SVD) for dimensionality reduction. - Cosine Similarity: Compute cosine similarity between book features.**

**4.5 Recommendation System Development:**

**While not explicitly model-based like classification, the focus is on recommending books using similarity metrics. Develop and test the function to recommend books based on the similarity scores from the SVD and TF-IDF.**

**4.7 Prediction:**

**Use the recommendation function to provide book suggestions based on a sample book title and evaluate its performance.**

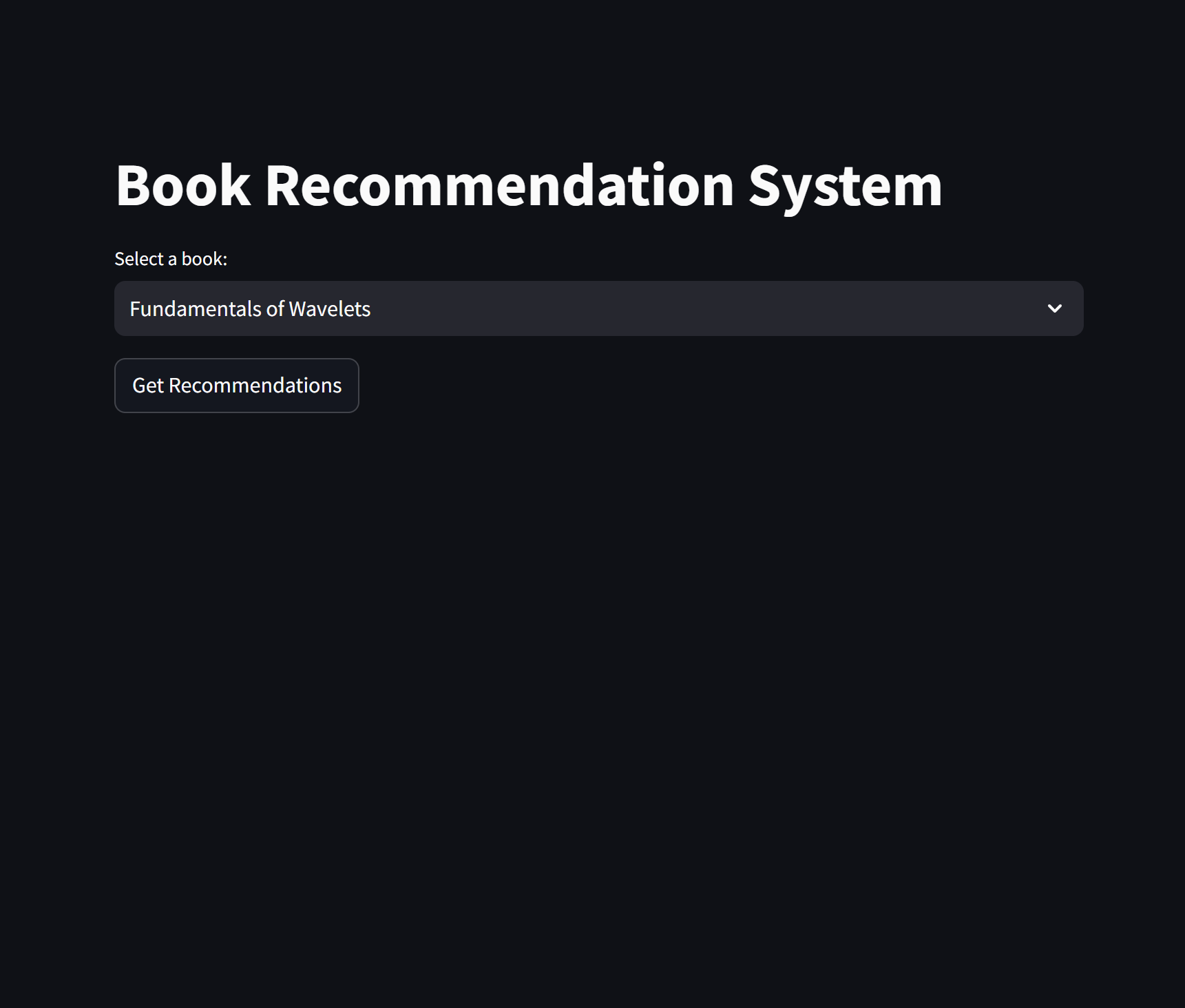
**4.8 Save model:**

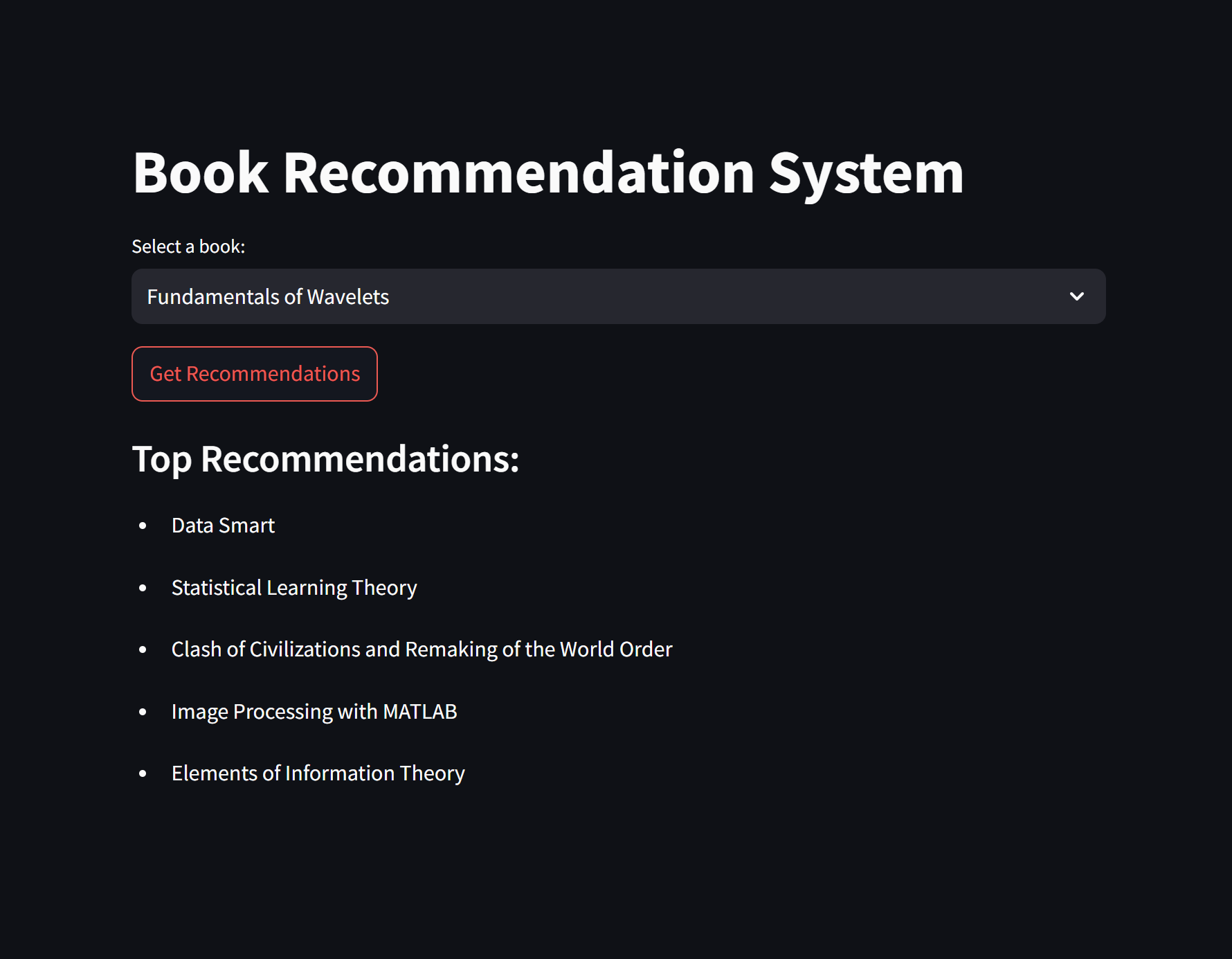
**Save the processed data and similarity matrix using pickle for deployment.**

**4.9 Deployment:**

**Deploy the Streamlit app on a cloud platform to make the recommendation system accessible online.**

**Here are the images of Application:**

****

****