

Architecture

Book Recommender System

Version 1.0

Date of Revision 15 April 2022

Contents

Document Version Control

1. Introduction

1.1 What is an architecture design document?

1.2 Scope

1.3 Constraints

2. Technical specifications

2.1 Dataset

2.1.1 Dataset overview

2.1.2 Input schema

2.2 User Input Requirements

2.3 Logging

2.4 Deployment

3. Technology Stack

4. Proposed Solution

5. Data pre-processing workflow

6. Admin data setup workflow

7. Machine learning training workflow

8. Machine learning recommendation workflow

9. Test cases

10. Performance

Document Version Control

Date Issued	Version	Description	Author
27th February 2022	1	Initial Architecture – V1.0	Savan Javia

1. Introduction

1.1 What is an architecture design document?

The architecture design document is a technical document describing the components and specifications required to support the solution and ensure that the specific business and technical requirements of the design are satisfied. The architecture design document is a technical document describing the components and specifications required to support the solution and ensure that the specific business and technical requirements of the design are satisfied.

1.2 Scope

This Book Recommender System web application will recommend books based on the user's interested book.

1.3 Constraints

The Book Recommender System application must be user friendly, as automated as possible and users should not be required to know any of the workings.

2. Technical specifications

2.1 Dataset

```
In [3]: df
```

```
Out[3]:
```

	Title	Author	Genre	SubGenre	Height	Publisher
0	Fundamentals of Wavelets	Goswami, Jaideva	tech	signal_processing	228	Wiley
1	Data Smart	Foreman, John	tech	data_science	235	Wiley
2	God Created the Integers	Hawking, Stephen	tech	mathematics	197	Penguin
3	Superfreakonomics	Dubner, Stephen	science	economics	179	HarperCollins
4	Orientalism	Said, Edward	nonfiction	history	197	Penguin
...
206	Structure and Randomness	Tao, Terence	science	mathematics	252	NaN
207	Image Processing with MATLAB	Eddins, Steve	tech	signal_processing	241	NaN
208	Animal Farm	Orwell, George	fiction	classic	180	NaN
209	Idiot, The	Dostoevsky, Fyodor	fiction	classic	197	NaN
210	Christmas Carol, A	Dickens, Charles	fiction	classic	196	NaN

211 rows × 6 columns

2.1.1 Dataset overview

- **Title:** contains the book title
- **Author:** contains the book Author
- **Genre:** represent the category of interest
- **Subgenre:** Represent the sub-category of interest
- **Publisher:** contains publisher of the book

2.1.2 Input schema

Feature name	Null / Required
Title	Required
Author	
Genre	Required
Subgenre	Required
Publisher	

2.2 User Input Requirements

Users must input book titles, and the system will recommend similar books.

2.3 Logging

We should be able to log every activity done by the user.

- The System identifies at what step logging required
- The System should be able to log each and every system flow.
- Developers can choose logging methods. You can choose database logging/ File logging as well.
- System should not be hung even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

2.4 Deployment



3. Technology Stack

Frontend	Html, CSS, Javascript, JQuery
Backend	Python, Flask
Database	Cassandra
Machine Learning	SK Learn
Deployment	Heroku

4. Proposed Solution

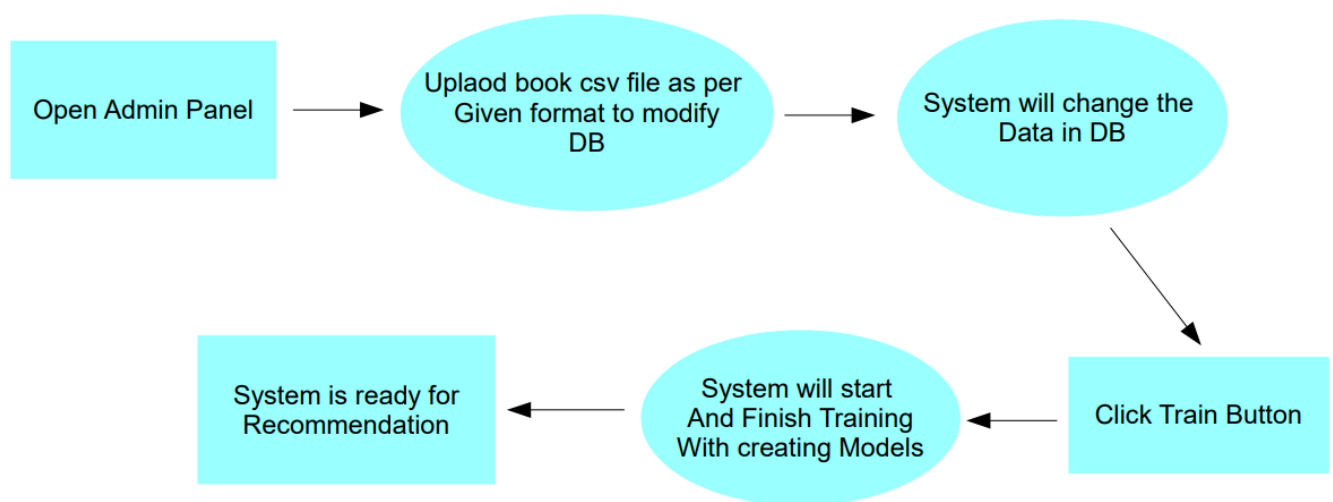
The proposed solution for this project is a Machine learning model that can recommend books to users based on their interest. Here we use a box of words text vectorization model to find similarity along with title, genre, subgenre, author, publisher.

5. Data pre-processing workflow

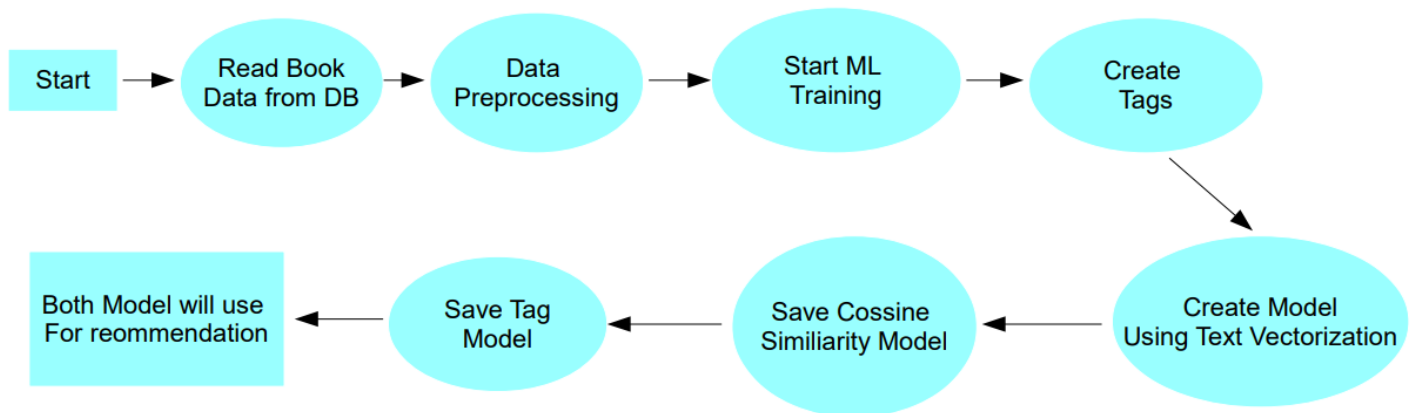
In data pre-processing the following steps are taken.

- **Create tags:** by combining feature columns we are creating tags.
- **Apply Stemming:** in tags we are applying stemming so the model will learn reliably. Ex: applying stepping on `dancing`, `danced` will become `danc`
- **Remove Stop Words:** in tags we are removing stop words like `the`, `and` etc... so, learning accuracy will increase.

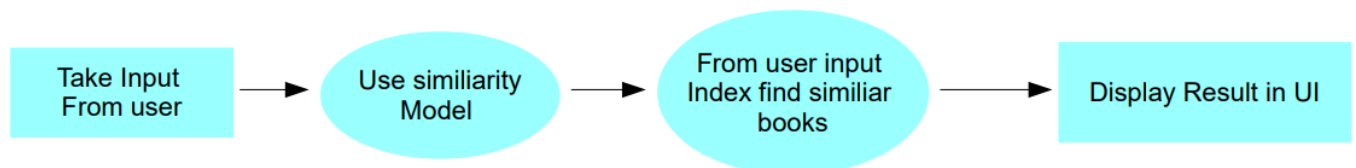
6. Admin data setup workflow



7. Machine learning training workflow



8. Machine learning recommendation workflow



9. Test cases

Sr No	User Input	Output	Test Result
1	Statistical Learning Theory	1. Nature of Statistical Learning Theory 2. Machine Learning for Hackers 3. Statistical Decision Theory' 4. Pattern Classification 5. Neural Networks	PASS

10. Performance

After the user will click `Recommend!` system takes less than a second to recommend + transferring result + displaying result on heroku.