LOW LEVEL DOCUMENT (BRS)

Contents

1. Introduction

1.1. What is Low-Level design document?

1.2. Scope

2. Architecture

3. Architecture Description

3.1. Data Description

3.2. Web Scrapping

3.3. Data Transformation

3.4. Data Insertion into Database

3.5. Export Data from Database

3.6. Data Pre-processing

3.7. Data Clustering

3.10. Model Building

3.11. Data from User

3.12. Data Validation

3.13. User Data Inserting into Database

3.14. Data Clustering

3.15. Model Call for Specific Cluster

3.16. Deployment

4. Unit Test Cases

1. Introduction

1.1. What is Low-Level design document?

The goal of LLD or a low-level design document (LLDD) is to give the internal logical design of the actual program code for Book Recommendation System. LLD describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document.

1.2. Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work

2. Architecture

3. Architecture Description

3.1. Data Description

We have 3 files in our dataset which is extracted from some books selling websites.

- Books – first are about books which contain all the information related to books like an author, title, publication year, etc.

- Users – The second file contains registered user's information like user id, location.
- Ratings – Ratings contain information like which user has given how much rating to which book.

So based on all these three files we can build a powerful collaborative filtering model. let's get started.

3.2. Web Scrapping

Web scraping (or data scraping) is a technique used to collect content and data from the internet. This data is usually saved in a local file so that it can be manipulated and analyzed as needed. If you've ever copied and pasted content from a website into an Excel spreadsheet, this is essentially what web scraping is, but on a very small scale.

all web scraping bots follow three basic principles:

- Step 1: Making an HTTP request to a server
- Step 2: Extracting and parsing (or breaking down) the website's code
- Step 3: Saving the relevant data locally

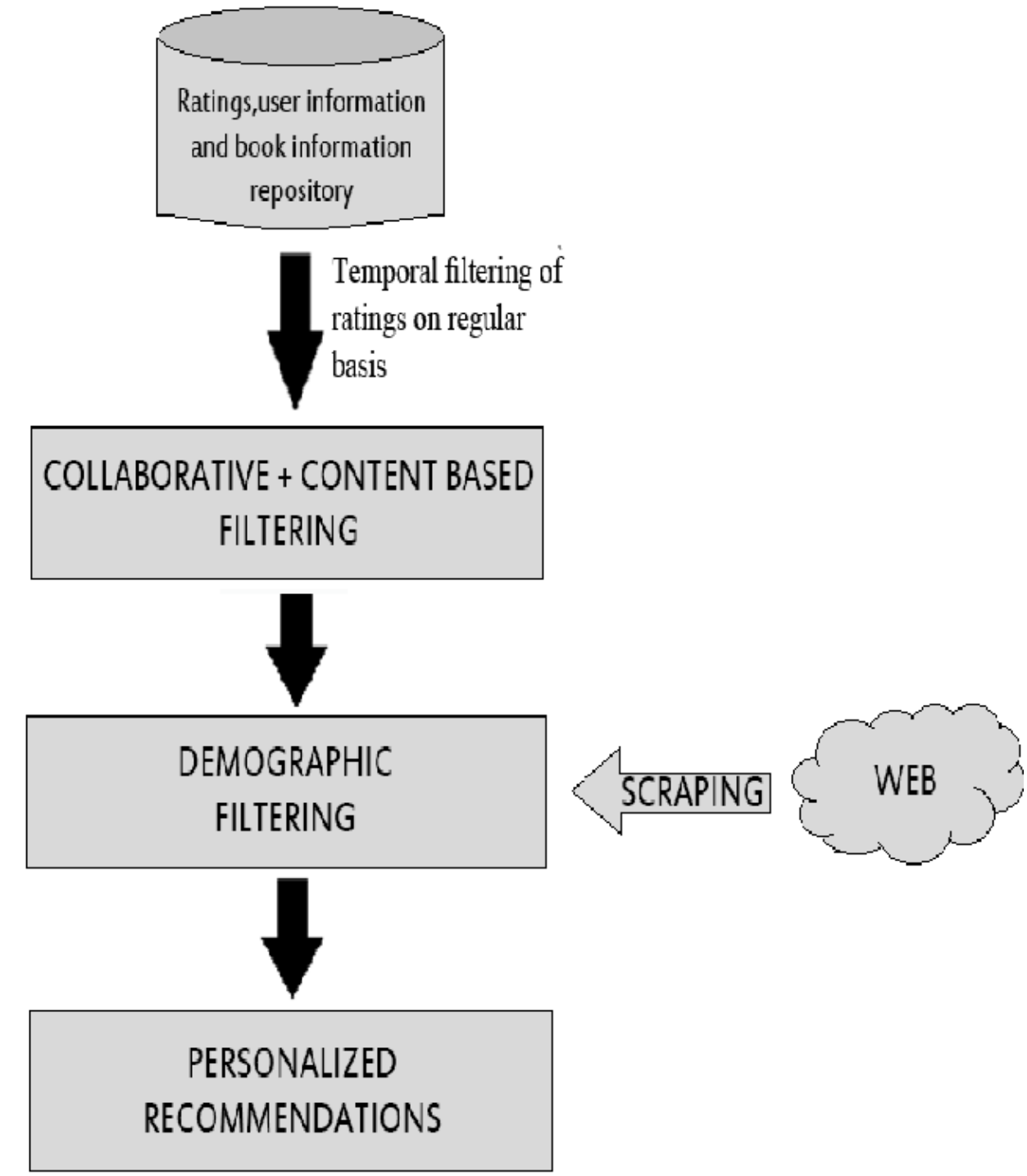


Fig. 2.Workflow of the Recommendation Process

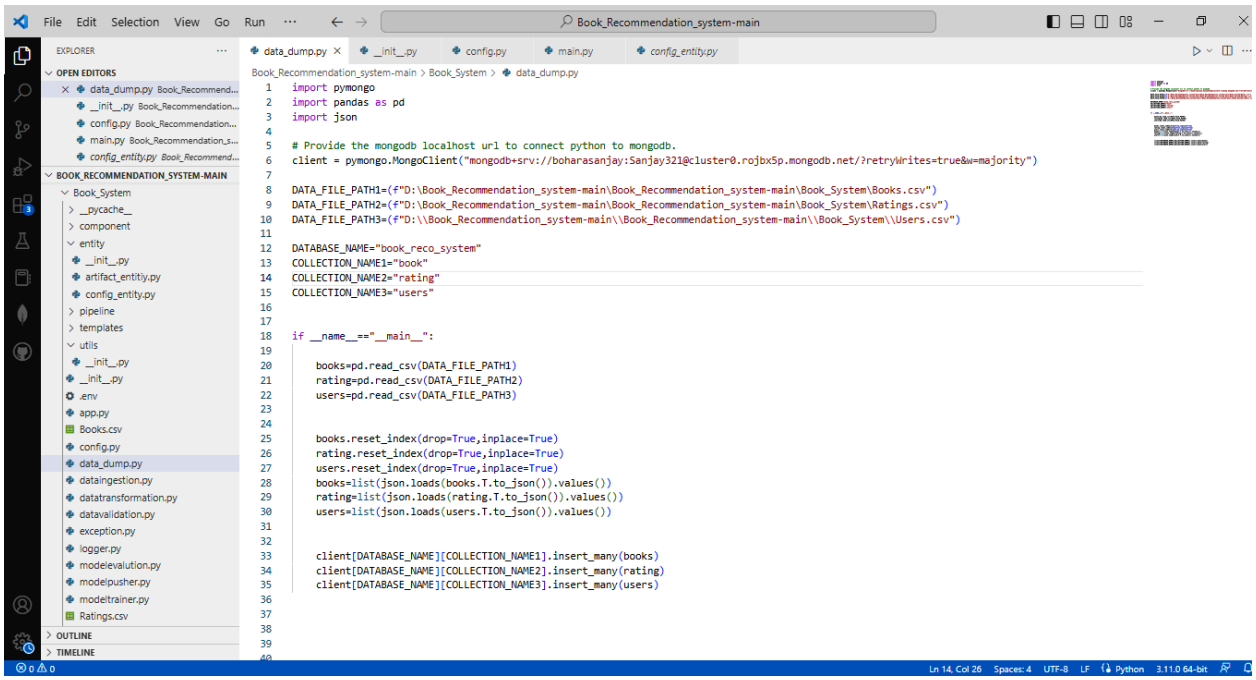
3.3. Data Transformation

Let us start while importing libraries and load datasets. While loading the file we have some problems like.

- The values in the CSV file are separated by semicolons, not by a comma.
- There are some lines which not work like we cannot import it with pandas and It throws an error because python is Interpreted language.
- Encoding of a file is in Latin

So while loading data we have to handle these exceptions and after running the below code you will get some warning and it will show which lines have an error that we have skipped while loading.

3.4. Data Insertion into Database



```python
import pymongo
import pandas as pd
import json

# Provide the mongodb localhost url to connect python to mongodb.
client = pymongo.MongoClient("mongodb+srv://boharasanjay:Sanjay321@cluster0.rojbx5p.mongodb.net/?retryWrites=true&w=majority")

DATA_FILE_PATH1=(f"D:\Book_Recommendation_system-main\Book_Recommendation_system-main\Book_System\Books.csv")
DATA_FILE_PATH2=(f"D:\Book_Recommendation_system-main\Book_Recommendation_system-main\Book_System\Ratings.csv")
DATA_FILE_PATH3=(f"D:\\Book_Recommendation_system-main\\Book_Recommendation_system-main\\Book_System\\Users.csv")

DATABASE_NAME="book_reco_system"
COLLECTION_NAME1="book"
COLLECTION_NAME2="rating"
COLLECTION_NAME3="users"


if __name__=="__main__":

    books=pd.read_csv(DATA_FILE_PATH1)
    rating=pd.read_csv(DATA_FILE_PATH2)
    users=pd.read_csv(DATA_FILE_PATH3)


    books.reset_index(drop=True,inplace=True)
    rating.reset_index(drop=True,inplace=True)
    users.reset_index(drop=True,inplace=True)
    books=list(json.loads(books.T.to_json()).values())
    rating=list(json.loads(rating.T.to_json()).values())
    users=list(json.loads(users.T.to_json()).values())


    client[DATABASE_NAME][COLLECTION_NAME1].insert_many(books)
    client[DATABASE_NAME][COLLECTION_NAME2].insert_many(rating)
    client[DATABASE_NAME][COLLECTION_NAME3].insert_many(users)
```

3.5. Export Data from Database

Data Export from Database - The data in a stored database is exported as a CSV file to be used for Data Pre-processing and Model Training.

3.6. Data Pre-processing

Now in the books file, we have some extra columns which are not required for our task like image URLs. And we will rename the columns of each file as the name of the column contains space, and uppercase letters so we will correct as to make it easy to use.

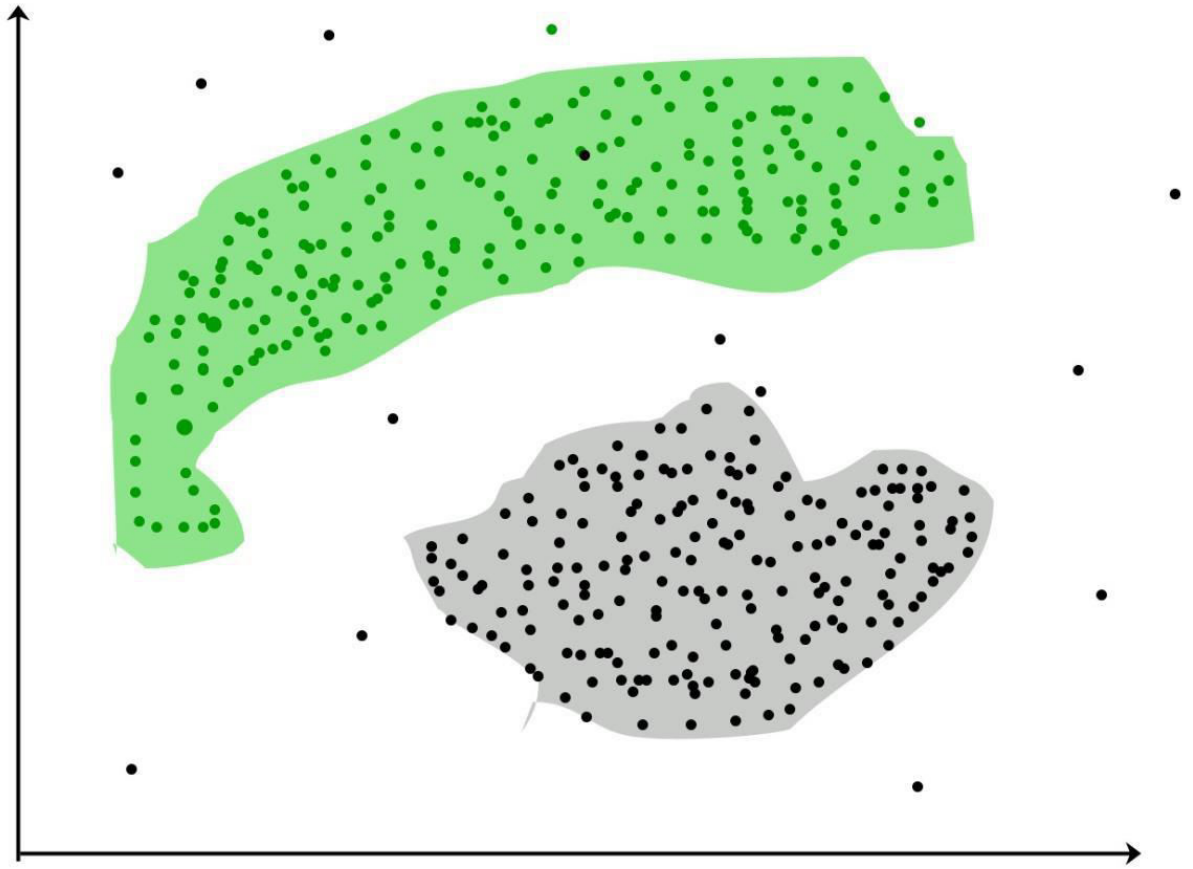
books = books[['ISBN', 'Book-Title', 'Book-Author', 'Year-Of-Publication', 'Publisher']]

```
books.rename(columns = {'Book-Title':'title', 'Book-Author':'author', 'Year-Of-Publication':'year', 'Publisher':'publisher'}, inplace=True)
users.rename(columns = {'User-ID':'user_id', 'Location':'location', 'Age':'age'}, inplace=True)
ratings.rename(columns = {'User-ID':'user_id', 'Book-Rating':'rating'}, inplace=True)
```

3.7. Data Clustering

Clustering is an unsupervised learning method in which we draw references from datasets consisting of input data without labelled responses. Generally, it is used as a process to find meaningful structure, explanatory underlying processes, generative features, and groupings inherent.

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them. Clustering is very important as it determines the intrinsic grouping among the unlabeled data present. There are no criteria for good clustering. It depends on the user, what is the criteria they may use which satisfy their need. This algorithm must make some assumptions which constitute the similarity of points and each assumption make different and equally valid clusters.



3.10. Model Building

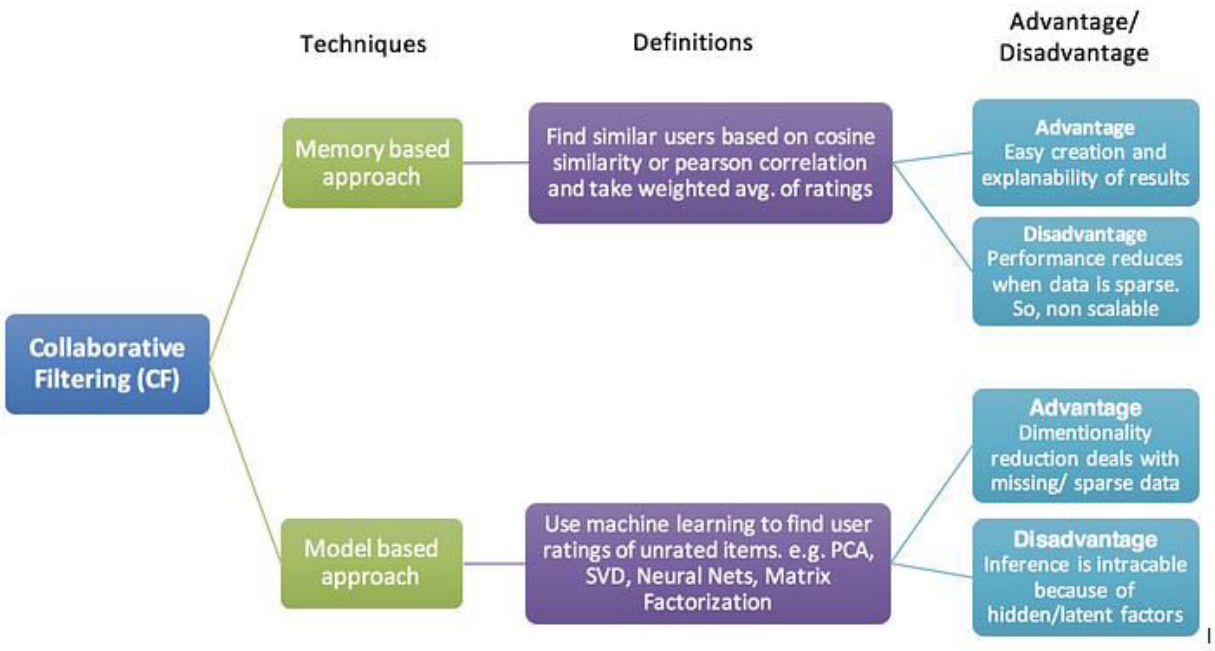
The obvious step in building a recommendation engine is finding the best-rated books that are a must-read for everyone. To do it, you have to get all book ratings and calculate the average rating score of every book in the dataset. After each book has a rating, the best books are easily extracted

3.11. Data from User

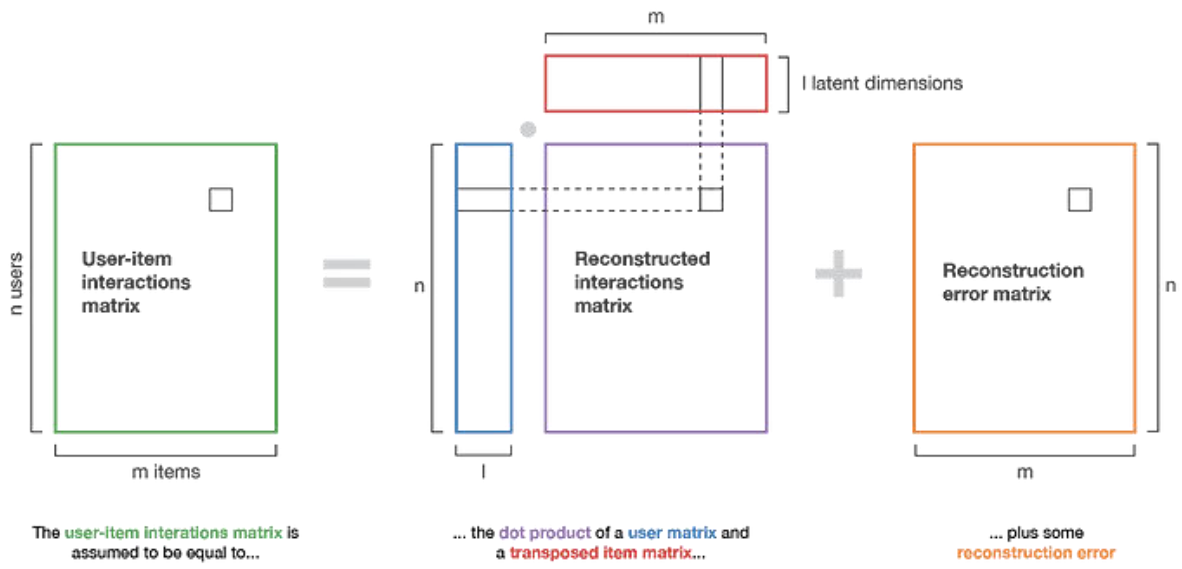
We have built a machine learning model for recommending books and now we will need to create a function using Python. When this function is called, we will have to pass the name of the book to it. The model will try to find books based on the features. We'll store those book names that the system recommends in a list and return them at the end.

3.12. Data Validation

We do not want to find a similarity between users or books. we want to do that If there is user A who has read and liked x and y books, And user B has also liked this two books and now user A has read and liked some z book which is not read by B so we have to recommend z book to user B. This is what collaborative filtering is.



So this is achieved using Matrix Factorization, we will create one matrix where columns will be users and indexes will be books and value will be rating. Like we have to create a Pivot table.



3.13. User Data Inserting into Database

1. Database Creation and connection - Create a database with name passed. If the database is already created, open the connection to the database.

2. Table creation in the database.

3. Insertion of files in the table

3.14. Data Clustering

Clustering Methods:

 Density-Based Methods: These methods consider the clusters as the dense region having some similarity and different from the lower dense region of the space. These methods have good accuracy and ability to merge two clusters.

 Example: DBSCAN (Density-Based Spatial Clustering of Applications with Noise), OPTICS (Ordering Points to Identify Clustering Structure) etc.

 Hierarchical Based Methods: The clusters formed in this method forms a tree-type structure based on the hierarchy. New clusters are formed using the previously formed one. It is divided into two categories:

o **Agglomerative** (*bottom up approach*)

o **Divisive** (*top down approach*)

 Examples: CURE (Clustering Using Representatives), BIRCH (Balanced Iterative Reducing Clustering and using Hierarchies) etc.

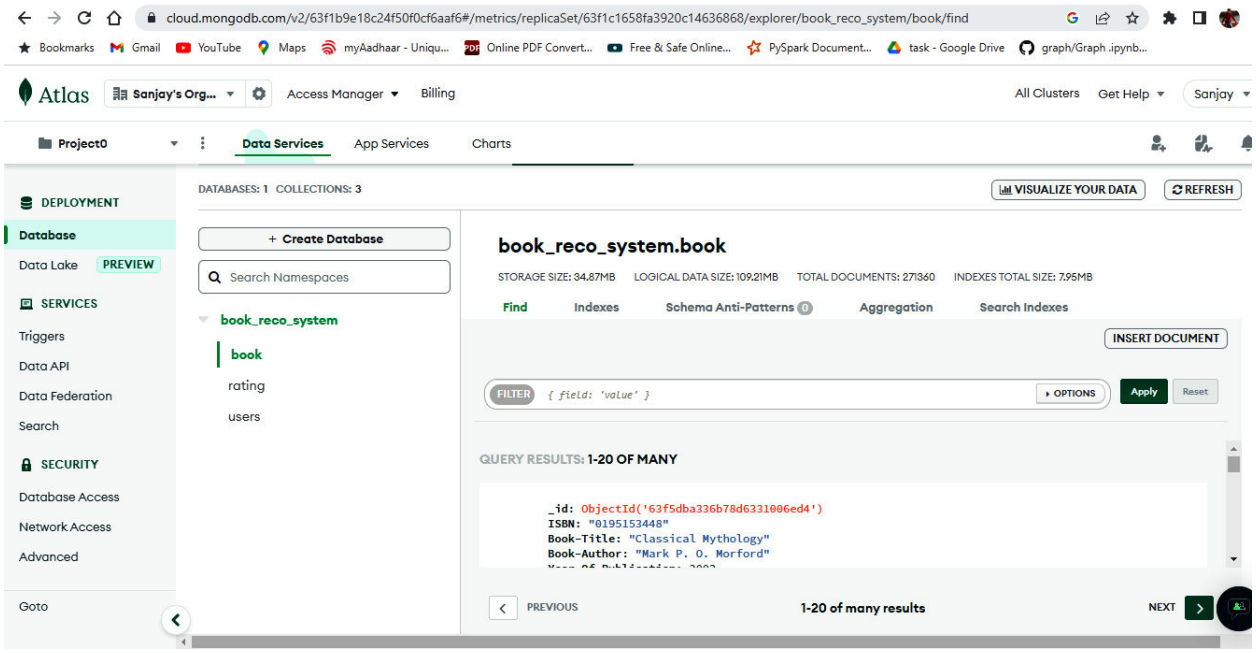
 Partitioning Methods: These methods partition the objects into k clusters and each partition forms one cluster. This method is used to optimize an objective criterion similarity function such as when the distance is a major parameter

example K-means, CLARANS (Clustering Large Applications based upon Randomized Search) etc.

 Grid-based Methods: In this method the data space is formulated into a finite number of cells that form a grid-like structure. All the clustering operation done on these grids are fast and independent of the number of data objects example STING (Statistical Information Grid), wave cluster, CLIQUE (Clustering In Quest) etc.

In this paper, partitioning method of clustering is used. We used Clustering algorithm which is simplest unsupervised learning algorithm in this paper and it partition n observations into k clusters where each observation belongs to the cluster.

3.15. Model Call for Specific Cluster

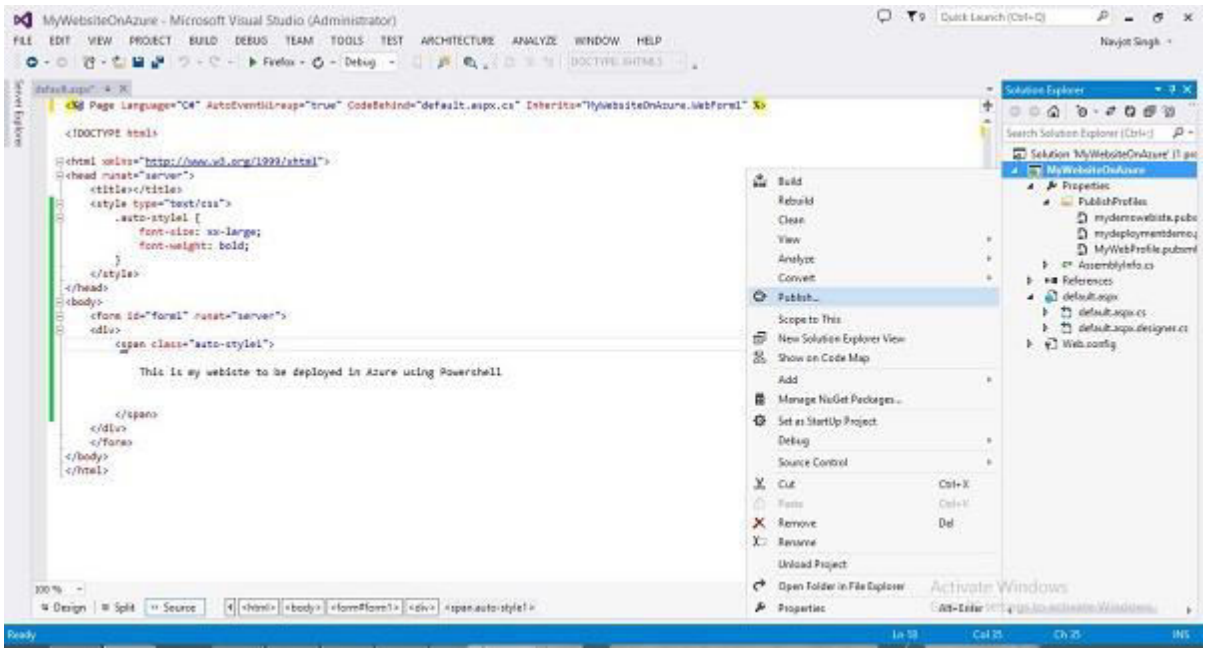


3.16. Deployment

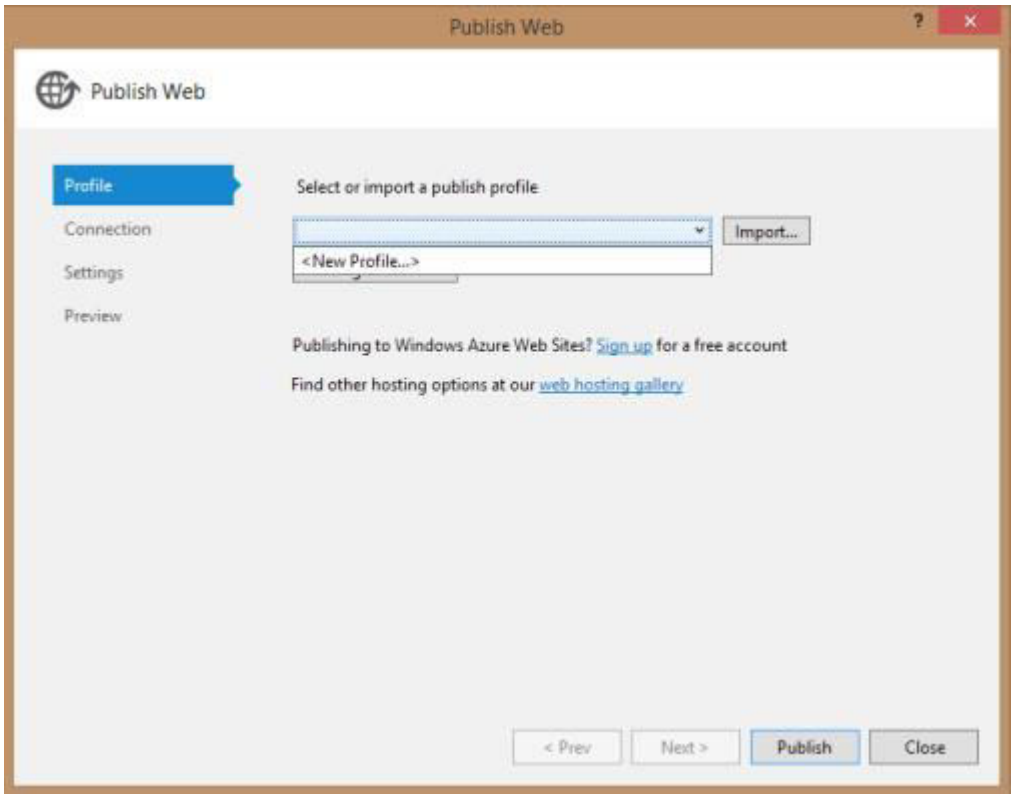
Create a Deployment Package

Step 1 − Go to your website in Visual Studio.

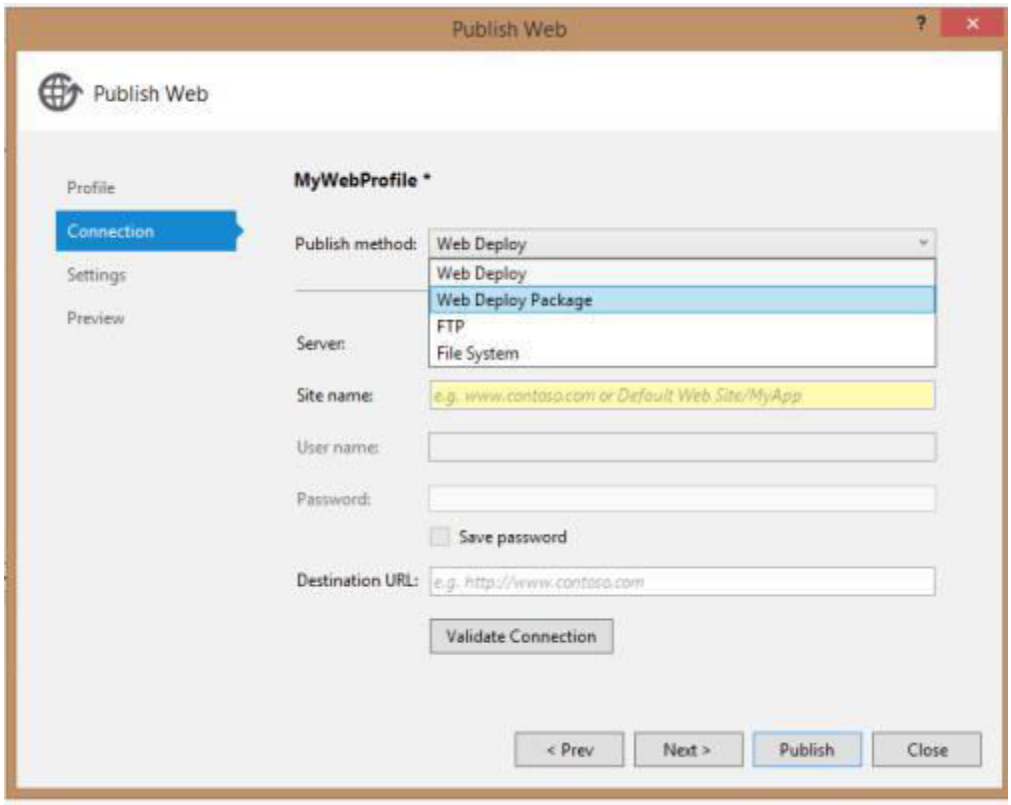
Step 2 − Right-click on the name of the application in the solution explorer. Select ‘Publish’.



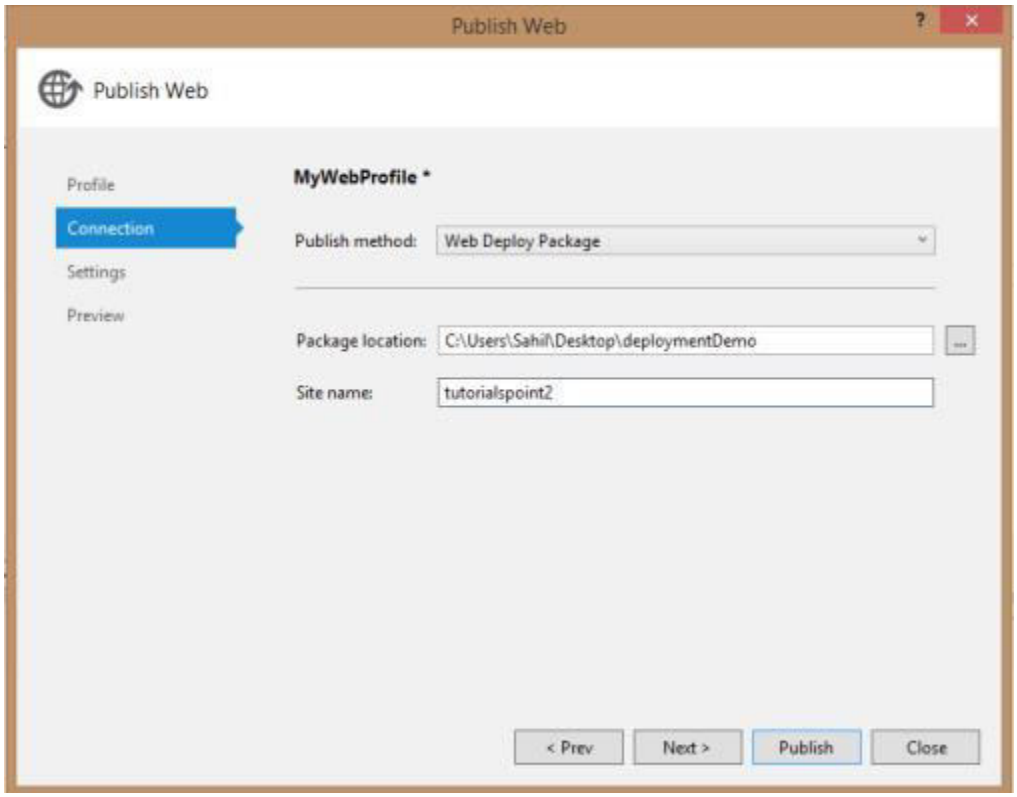
Step 3 − Create a new profile by selecting 'New Profile' from the dropdown. Enter the name of the profile. There might be different options in dropdown depending on if the websites are published before from the same computer.



Step 4 − On the next screen, choose ‘Web Deploy Package’ in Publish Method.



Step 5 − Choose a path to store the deployment package. Enter the name of site and click Next.



Step 6 − On the next screen, leave the defaults on and select ‘publish’.

After it’s done, inside the folder in your chosen location, you will find a zip file which is what you need during deployment.

4. Unit Test Cases

Test Case Description	Pre-Requisite	Expected Result
Verify whether the Application URL is accessible to the user	Application URL should be defined	Application URL should be accessible to the user
Verify whether the Application loads completely for the user when the URL is accessed	1.Application URL is accessible 2.Application is deployed	The Application should load completely for the user when the URL is accessed
Verify whether user gets Submit button to submit the inputs	Application is accessible	User should get Submit button to submit the inputs
Verify whether user is presented with recommended results on clicking submit	1. Application is accessible 2. User is signed up to the application 3. User is logged in to the application	User should be presented with recommended results on clicking submit
Verify whether the recommended results are in accordance to the selections user made	1. Application is accessible	The recommended results should be in accordance to the selections user made