Book Recommendation Engine Using KNN

Project Repo: [Books Recommendation Engine using KNN (github.com)]

Submitted By: B1 - G4	
Name	**Reg. No.**
Homaira Adiba	18201052
Md. Shahib Anower	18201055
Md. Maidul Islam	18201059

Submitted To
Fahad Ahmed
Lecturer, Dept. of Computer Science & Engineering
University of Asia Pacific (UAP)

Department of Computer Science & Engineering
Fall - 2021

List of Contents:

Motivation:

Recommender systems are a useful alternative to search algorithms since they help users discover items they might not have found otherwise. Of note, recommender systems are often implemented using search engines indexing non - traditional data. There are many approaches for recommender systems used in the development of the machines through– Content based approach and Collaborative approach or Hybrid approach (combining the above two approaches).

Various kinds of readers prefer various kinds of books or genres. In order to get book recommendations efficiently or decide what to read next, a good book recommendation engine would be helpful for the existing readers in our society.

Objectives:

Recommendation systems are used for the purpose of suggesting items to purchase or to see. They direct users towards those items which can meet their needs by cutting down a large database of Information. Various techniques have been introduced for recommending items i.e., content, collaborative and association mining techniques are used.

This project helps in solving the problem of data sparsity using the KNN (K-Nearest Neighbor) algorithm to achieve better performance.

Readers can get better book recommendations if the project is implemented.

Problem Statement:

Sometimes newbie readers or even veteran readers pick a book randomly which does not go with their preferred genre. It often leads to dissatisfaction with the reading experience and demotivates them to read more.

A recommender system helps people who do not have sufficient personal experience to evaluate the number of alternatives offered by a website. It provides consumer with information to help them decide which items to purchase. The proposed work is different from existing recommender systems since the existing

only considers the recommending the items based on user ratings of item. It doesn't recommend items when ratings for an item are not available. The proposed system uses combination of collaborative filtering and association mining. Collaborative filtering is used for finding similarity between items which would help the system to recommend items and association mining is used for filling the vacant ratings where necessary. Then it uses prediction of target user to the target item using item based collaborative filtering. Thus, the use of both methods can help to manage data sparsity problem and cold start problem in recommender system

Impact on Society:

The project will help to choose books accordingly which will ease the reading experience of the readers. A better reading experience will attract more people to reading. Software engineers can solve similar problems or test the accuracy of similar problem solving using KNN and give validity if it is reasonable or not to give recommendations using KNN algorithm.

This recommendation is provided to the users by developing a decision machine which accepts the training datasets and showcasing the resulted outputs to the users as generated by the machine. The expanding requests of online Data have led to imagine new systems for organizing and showing things of client interests. This venture utilizes thing-based community sifting method to create appraisals. The Thing based synergistic separating can evacuate the information sparsity issue and can give great proposal. At last, the consequences of comparability figuring give great execution at exactness Since, we are looking at different techniques in this undertaking it encourages us in contrasting the time unpredictability of the considerable number of strategies and aides in discovering which strategy is increasingly responsive in less interim and gives progressively suitable outcomes. In the future work, there are many different methods which are used in mining the data and can be used for recommendation process.

Related works and Background Study:

1. Machine learning algorithm. (KNN)

   In statistics, the k-nearest neighbors' algorithm is a non-parametric supervised learning method first developed by Evelyn Fix and Joseph Hodges in 1951, and later expanded by Thomas Cover. It is used for classification and regression. In both cases, the input consists of the k closest training examples in a data set.

2. *Streamlit Framework.*

   Streamlit is **an open-source app framework in Python language**. It helps us create web apps for data science and machine learning in a short time. It is compatible with major Python libraries such as scikit-learn, Keras, PyTorch, SymPy(latex), NumPy, pandas, Matplotlib etc.

   Official Streamlit Documentation.

We have found some related work to our research. They are:

[1] Proposed a general framework for content-boosted collaborative filtering. This work improved on recommendation by boosting collaborative filtering algorithms with contents. Hence with their results, it was clear that a naïve hybrid method tends to perform better than a pure content-based or collaborative filtering algorithm.

[2] Carried out a survey of collaborative filtering techniques by first identifying the collaborative filtering tasks and the challenges facing them such as data sparsest, scalability, gray sheep, shilling attacks, privacy protections etc. and possible solutions to these challenges. The different collaborative filtering techniques such as memory-based, model-based and hybrid collaborative filtering algorithms were represented and analyzed to determine their predictive performance and their ability to address the previously stated challenges.

[3] Proposed Amazon.com recommendations: an item-toitem collaborative filtering. These approaches solved recommendation problems with the use of an algorithm different from the traditional collaborative filtering algorithm, cluster models algorithm and search-based algorithm methods. The item-to-item

collaborative filtering algorithm applied to Amazon's online shop computations was able to scale independently of the number of customer's and items in the product catalog.

Solution Methodology:

- **Block Diagram:**

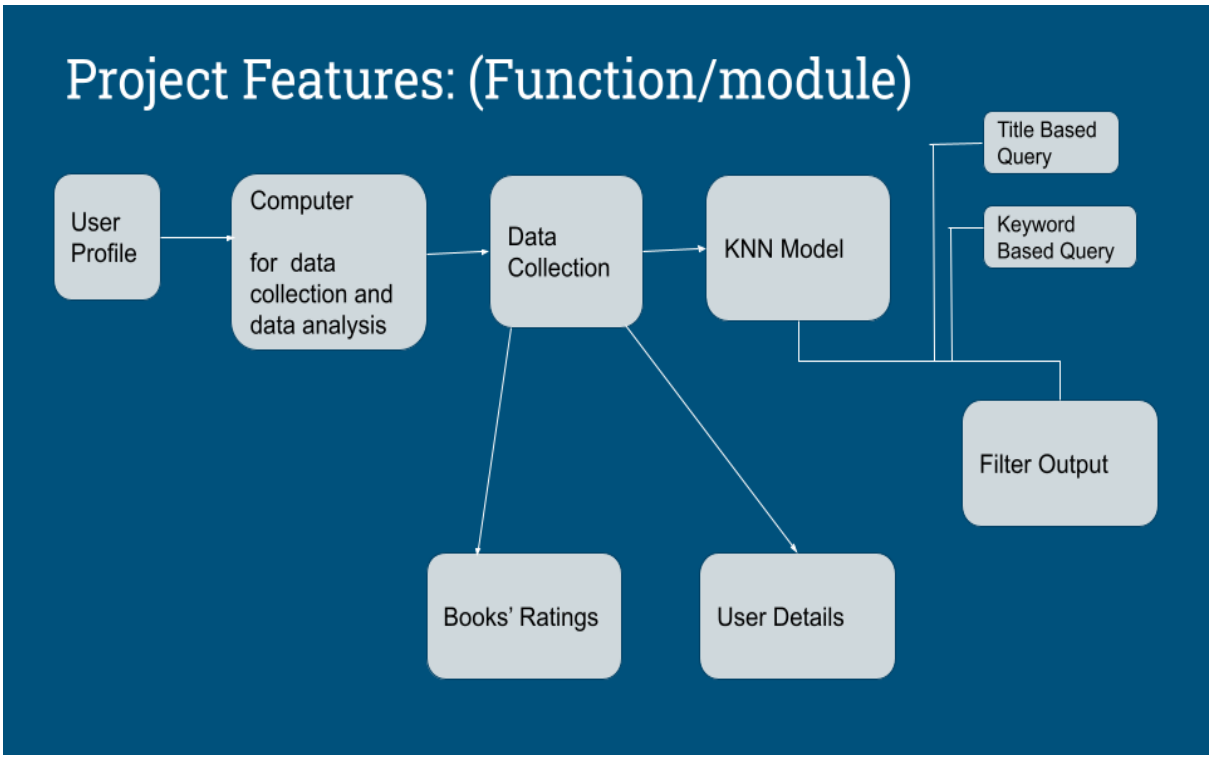
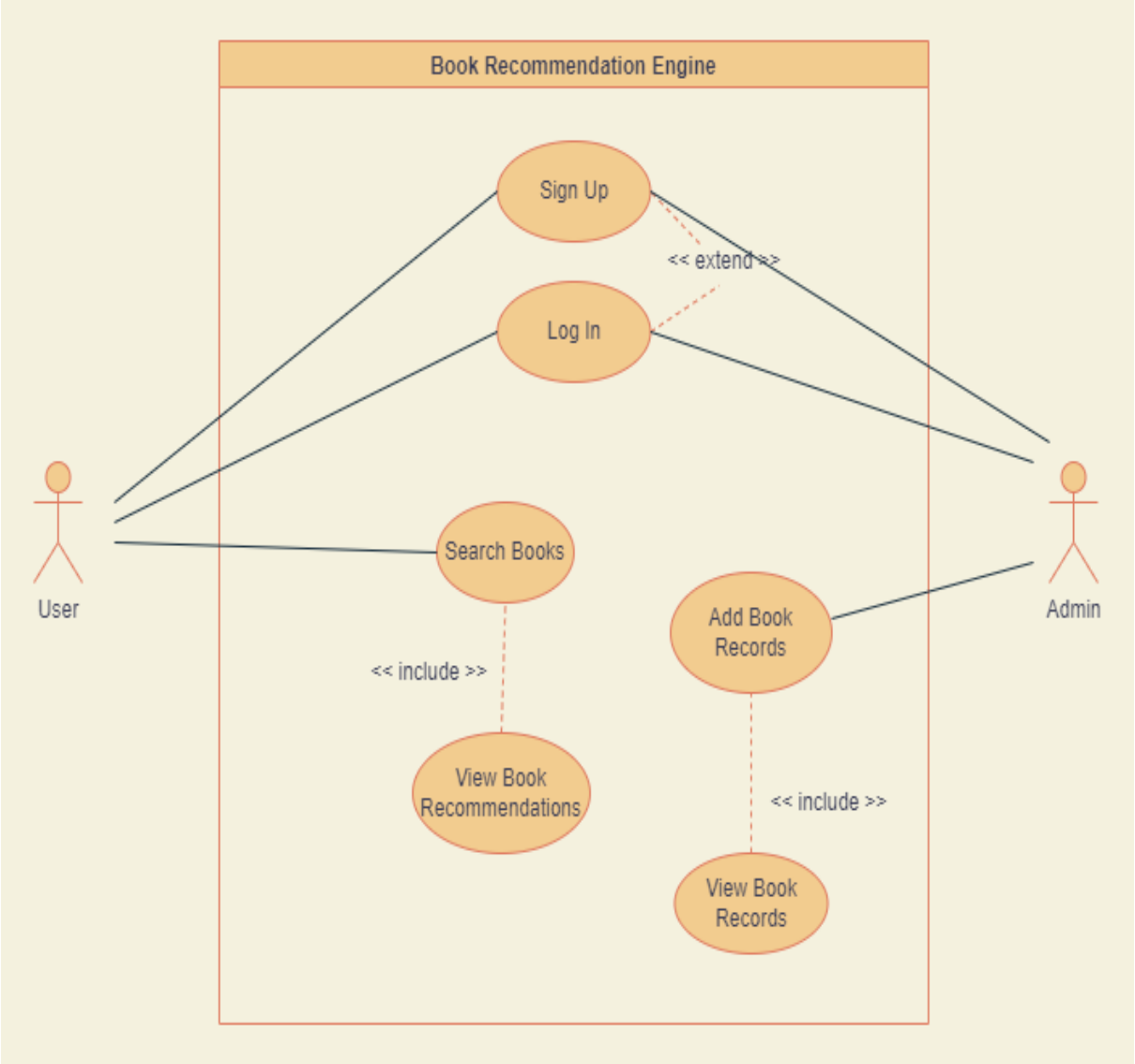


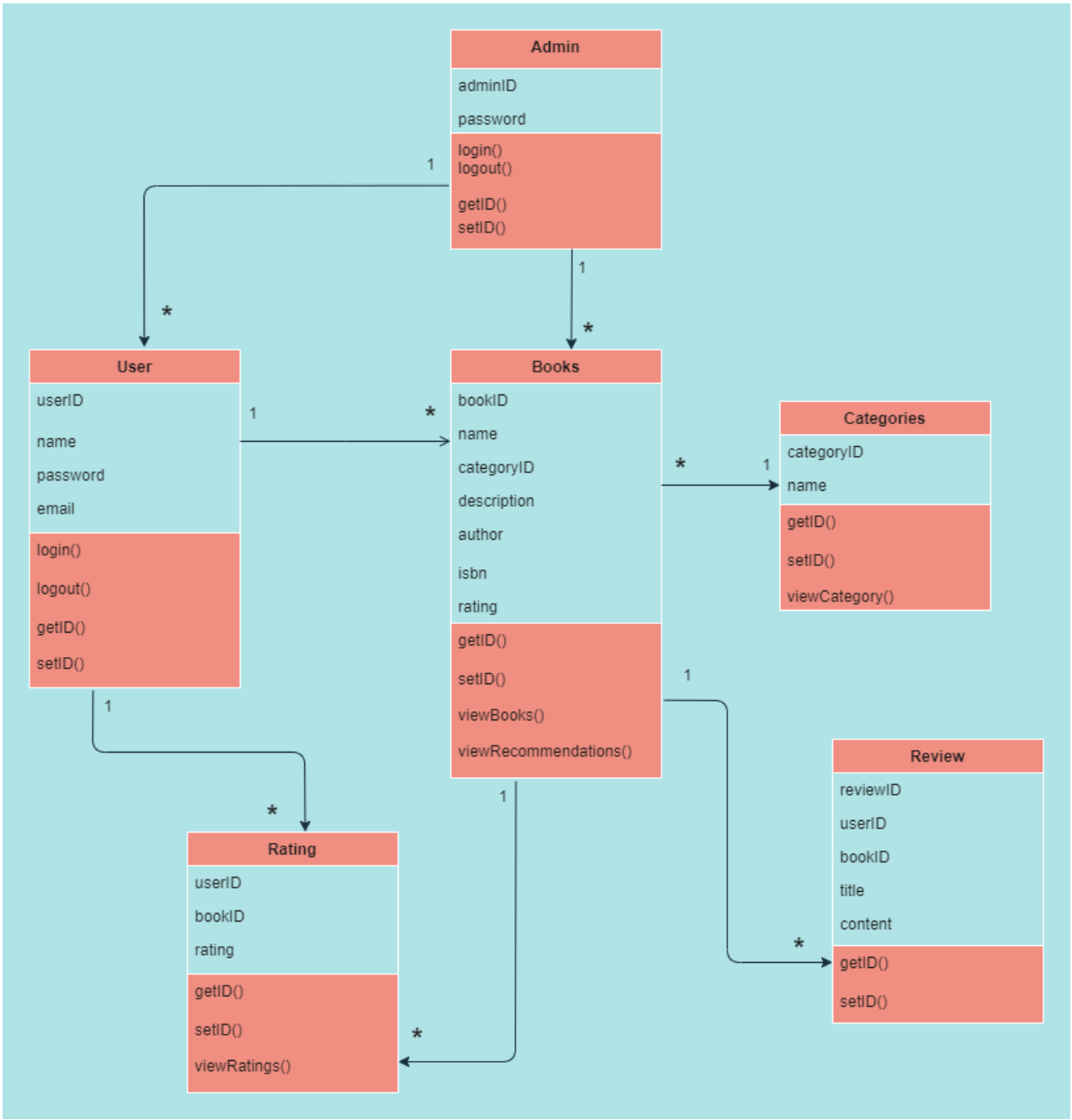
Fig.: Block Diagram of Book Recommendation using KNN

- **Use-Case-Diagram:**



- **Fig.: Use-Case-Diagram of Book Recommendation using KNN**

- **UML:**



- **Fig.: UML-Diagram of Book Recommendation using KNN**

Risk Analysis (Critical Challenges):

After working and analyzing it we realized that we are going to face various challenges. Those critical challenges can be:

- Accuracy– The accuracy of the output should be maximized. Inaccuracy will lead to wrong kinds of book recommendations.

- Algorithm – Among all the available algorithms, K-Nearest Neighbor Algorithm is chosen and the suitability of this algorithm should be tested.

Troubleshooted every issue from the official documentations of several packages.

Result (Final Output of the Project):

Dataset Used: [Book-Crossing-Dataset]

Script: [colab-note]

This code:

```
get_recommends("The Queen of the Damned (Vampire Chronicles (Paperback))")
```

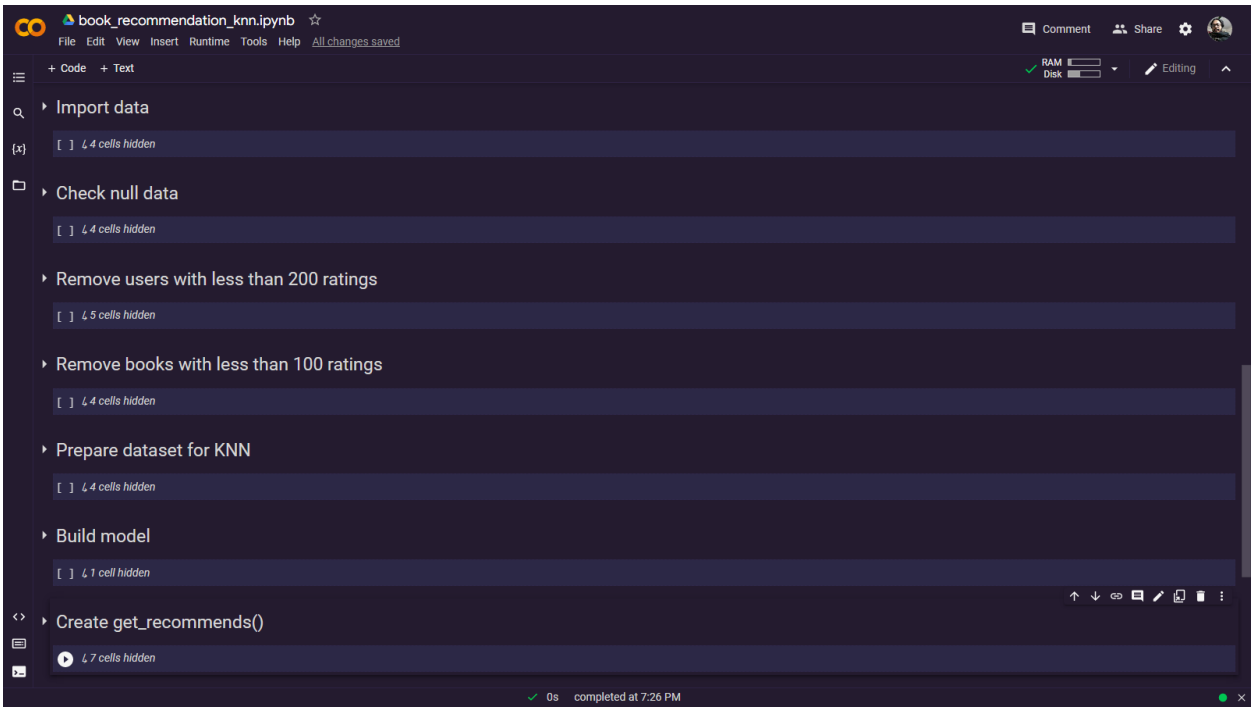
should return:

```
[
  'The Queen of the Damned (Vampire Chronicles (Paperback))',
  [
    ['Catch 22', 0.793983519077301],
    ['The Witching Hour (Lives of the Mayfair Witches)', 0.7448656558990479],
    ['Interview with the Vampire', 0.7345068454742432],
    ['The Tale of the Body Thief (Vampire Chronicles (Paperback))', 0.5376338362693787],
    ['The Vampire Lestat (Vampire Chronicles, Book II)', 0.5178412199020386]
  ]
]
```

One more test with 'Animal Farm' is shown below.

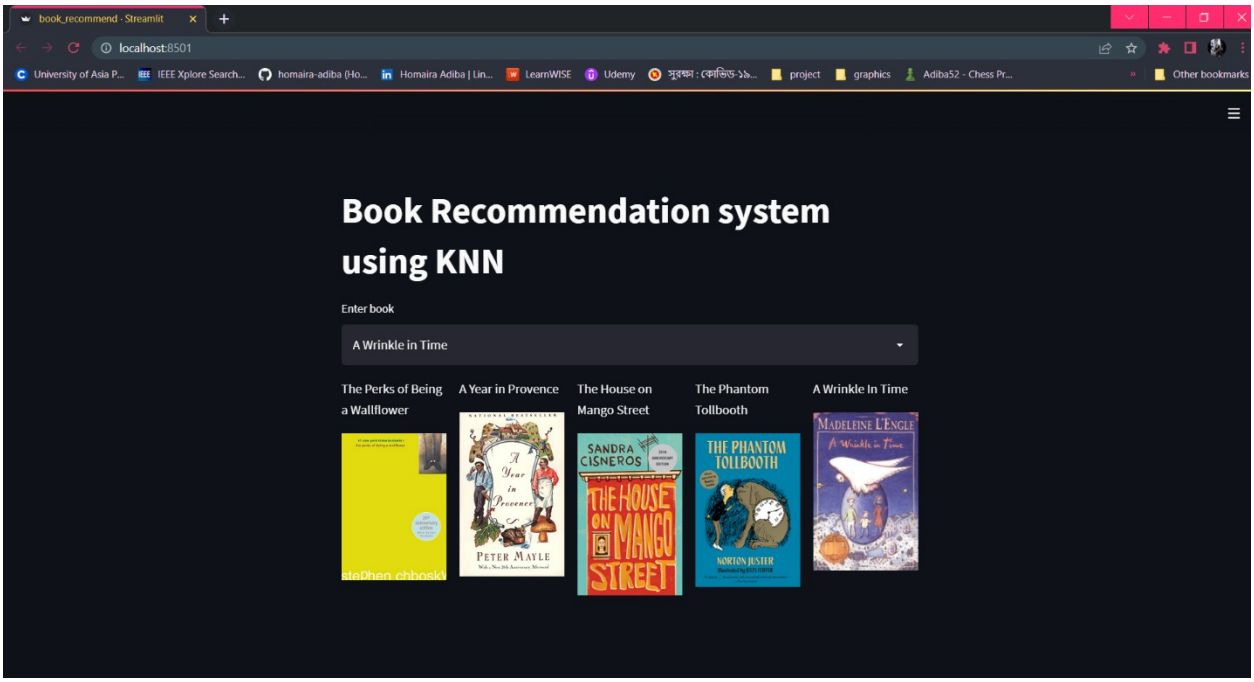
```
[31] get_recommends("Animal Farm")

     ['Animal Farm', array([['Winter Moon', 0.787234365940094],
            ['Call of the Wild', 0.7848600149154663],
            ['The Corrections: A Novel', 0.7625773549079895],
            ['1984', 0.7618247270584106],
            ['On the Road', 0.7497414350509644]], dtype=object)]
```

After Loading the Libraries, Datasets are imported. Then the dataset was checked is there is any null data. After that, users with less than 200 ratings and books with less than 100 ratings were removed in order to prepare the dataset for KNN and build the model. The after implementing get_recommends(title) function, the scripts was tested.



After that Project is implemented with UI to be more useable by users. It was done via Streamlit. A framework of Python which turn scripts into Web-Apps.

Here is a snapshot of the completed project:



Project Management:

Task	Duration (Weeks - Hours)	Cost (Hours*100) [BDT]
Project Proposal	1st to 3rd	(3*1000) = 3,000
Dataset Collection	4th to 6th	(5*1000) = 5,000
Design	7th to 9th	(10*1000) = 10,000
Build ML model	10th to 11th	(5*1000) = 5,000
Build the UI of the project	12th	(15*1000) = 15,000
Final Submission and Documentation	13th to 14th	(25*1000) = 25,000
		Total = 63,000

Learnings:

The increasing demands of Online Information have led to invent new techniques for prioritizing and presenting items of Users Interests. This project uses item-based Collaborative Filtering. To produce ratings. The Item based collaborative filtering can remove the data sparsity problem and can provide good recommendation. Finally, the results of similarity calculation give good performance at accuracy.

There are a few more things that can be done to build on this project further:

1. Implementing the project with better algorithms. This can be done using deep learning frameworks like TensorFlow.
2. Deploying this project on front-end frameworks like Django, Flask, Bokeh and Dash. All of these are python front-end frameworks. Cross-platform frameworks like R's shiny can also be used.

Our model made full use of the rating data to improve the accuracy of recommender systems.

References:

[1] Onah, D. and Sinclair, J. (2015). Collaborative filtering recommendation system: a framework in massive online courses. Proceedings of the 9th International Technology, Education and Development Conference, Madrid, Spain, 1249-1257.

[2] Su, X., and Khoshgoftaar T. M. (2009). A survey of collaborative filtering techniques. Advances in artificial intelligence. Hindawi publishing corporation. 1-19.

[3] Linden, G., Smith, B., and York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 7(1), 76–80.

Appendix A:

CEP Mapping:

How Ks are addressed through the project and mapping among Ks, COs, and POs

Ks	Attributes	How Ks are addressed through the project	COs	POs
K2	Mathematics	Conceptually based mathematics, numerical analysis, statistics and the formal aspects of computer and information science to support analysis and modeling applicable to the discipline.	CO1 CO2	PO-a PO-b
K3	Engineering Fundamentals	A systematic, theory-based formulation of engineering fundamentals required in the engineering discipline.	CO1 CO2	PO-a PO-b
K4	Engineering Specialization	Engineering specialist knowledge that provides theoretical frameworks and bodies of knowledge for the accepted practice areas in the engineering discipline; much is at the forefront of the discipline.	CO1 CO2	PO-a PO-b
K5	Engineering Design	The project supports engineering design practice all along.	CO3	PO-c
K8	Research	Engagement with selected knowledge in the research literature of the discipline.	CO1	PO-d

How Ps are addressed through the project and mapping among Ps, COs, and POs

Ps	Attribute	How Ps are addressed through the project	COs	POs
P1	Depth of Knowledge Requirement	Requiring a rigorous study of all the existing related problems (K8) and high-level coding language skill (K3, K4), the project is made.	CO1 CO2 CO3 CO4 CO10	PO-a PO-b PO-c PO-e PO-l
P3	Depth of Analysis Requirement	Have no obvious solution as there exist different algorithms and requires abstract thinking, and originality in analysis to formulate suitable models.	CO1 CO2 CO10	PO-a PO-b PO-l
P6	Extend to Stakeholders	Involve diverse groups of stakeholders, here the readers, with widely varying needs.	CO8	PO-j
P7	Interdependence	Interdependent components such as packages i.e NumPy, panda and Google Colab or JupyterLab and so on are part of our project.	CO7 CO9	PO-i PO-k

How As are addressed through the project

As	Attribute	How As are addressed through the project	COs	POs
A1	Range of Resources	This project involved the use of diverse resources information (datasets) and technologies (Colab).	CO-8	PO-j
A2	Level of Interaction	Required resolution of significant problems arising from interactions from between wide-ranging or conflicting technical, engineering or other issues.	CO-8	PO-j
A4	Consequences for society and the environment	This project has significant consequences in a range of contexts, characterized by difficulty of prediction and mitigation.	CO-8	PO-j