Skill Development Laboratory

|  |  |  |  |
| --- | --- | --- | --- |
| **TITLE OF PROJECT** – Goodreads: Books analysis & similar books recommendation | | | |
| **Abstract :**  The approach of recommender systems are based on the ground of users history of choices, likings and reviews, each of which is interpreted in order to project the future choices of the user. In this study, a novel and strong recommender system for the books is proposed. A content based book recommendation application was developed which makes recommendations according to user's taste and choices.  The basic idea behind analysing the Goodreads dataset is to get a fair idea about the relationships between the multiple attributes a book might have, such as the aggregrate rating of each book, the most rated books and books with more pages, etc.  We've always conisdered the magical persona books seem to hold, and with this notebook, we step out on a journey to see what kind of books really drives people to read in this era of modern smart devices.  With such a vast, overwhelming number of factors, we'll go over such demographics:   * Does any relationship lie between average\_ratings and the total ratings given? * Where do majority of the books lie, in terms of ratings - Does reading a book really bring forth bias for the ratings? * Do number of pages make an impact on reading styles, ratings and popularity? * Can books be recommended based on ratings? Is that a factor which can work? | | | |
| **Objectives:**  1.To analyze Goodreads dataset to get a fair idea about the relationships between the multiple attributes a book might have, such as the aggregrate rating of each book, the most rated books and books with more pages, etc.  2.To recommend movies to users by using k-means clustring and KNN algorithm | | | |
| **Dataset** –   * Type of dataset- .CSV * Source- 1) <https://www.kaggle.com/jealousleopard/goodreadsbooks> * Attributes-(bookID,title,authors,average\_rating,isbn,isbn13,language\_code, * # num\_pages,ratings\_count,text\_reviews\_count) | | | |
| **Libraries and Framework/ Platform –**   * Dash (GUI) * Scikit learn * Pandas * Numpy * Matplotlib | | | |
| **Machine Learning Algorithm-**   * K-means Clustering * KNN Algorithm | | | |
| No | **Roll Number** | **Name of the Student** | **Class** |
| 1. | 305073 | Nitin Ajay Dixit | TE-1 |
| 2. | 305074 | Vasant Raju Donurkar | TE-1 |
| 3. | 305075 | Sameer Shah | TE-1 |
| **Name of Guide-** Mhetre Mam | | | |