Abstract:

Machine learning is a scientific study of statistical model and algorithms. In this research I will use the machine learning algorithms, K-NN and matrix factorization. In the books recommendations system BX books dataset is used.

As the amounts of online books are exponentially increasing due to COVID-19 pandemic, finding relevant books from a vast e-book space becomes a tremendous challenge for online users. Personal recommendation systems have been emerged to conduct effective search which mine related books based on user rating and interest. Most of these existing systems are user-based ratings where content-based and collaborativebased learning methods are used. These systems' irrationality is their rating technique, which counts the users who have already been unsubscribed from the services and no longer rate books.

This paper proposed an effective system for recommending books for online users that rated a book using the clustering method and then found a similarity of that book to suggest a new book. The proposed system used the K-means Cosine Distance function to measure distance and Cosine Similarity function to find Similarity between the book clusters. Sensitivity, Specificity, and F Score were calculated for ten different datasets. The average Specificity was higher than sensitivity, which means that the classifier could re-move boring books from the reader's list. Besides, a receiver operating characteristic curve was plotted to find a graphical view of the classifiers' accuracy. Most of the datasets were close to the ideal diagonal classifier line and far from the worst classifier line. The result concludes that recommendations, based on a particular book, are more accurately effective than a user-based recommendation system.

Keywords :

- Problem Statement
- Data Summary
- Data Pipeline
- Feature Engineering
- Data Cleaning
- EDA
- Preparing Data for Model
- Applying Model
- Conclusion
- Challenges
- Future Scope

Introduction :

A recommender system is a type of information filtering system which predicts the rating or preference that a user would give to an item. Recommender systems are sometimes referred to as recommendation systems. Recommendation system describes web applications that predict response to options.

We make use of a hybrid recommender system to power our recommendations. Here content-based filtering and collaborative filtering are two types of recommender systems that can be combined to form our hybrid system. A content-based recommender system aims to predict a user's characteristics or behavior based on the item's characteristics, which he or she responds positively to [3]. For example, if a user likes the first book of Harry Potter, and if the second book is similar to the first, then it recommends the second book to

the user. Collaborative filtering on the other hand does not need the features of the items to be given. A feature vector characterizes every user and item. It collects user feedback on different items and uses them for recommendations [3]. Here user ratings are used to determine the user or item similarity in collaborative filtering. If there is a high correlation of users rating the first Harry Potter book and the second Harry Potter book, then they are deemed to be similar. Our hybrid system uses both of these approaches for more efficiency.

Now-a-days anything can be done on foot, which may include listening to the news, attending meetings, and even taking notes (with voice dictation). Reading is the only thing you can't do while walking. One of the best things about audiobooks is that they allow individuals to fit more reading into their already hectic schedules. Listening to audiobooks allows you to read in the “inbetween” moments of your day when you would otherwise have to do something else.

Problem Statement:

- During the last few decades, recommender systems have taken more and more place in our lives. From e-commerce to online advertisement, recommender systems are today unavoidable in our daily online journeys.
- Recommender systems are really critical in some industries as they can generate a huge amount of income when they are efficient or also be a way to stand out significantly from competitors.
- Recommender systems are algorithms aimed at suggesting relevant items to users. The main objective is to create a book recommendation system for users.

Data Summary:

The dataset is comprised of three csv files:: User_data, Book_data, Rating_data

❖ Users_dataset.

User-ID (unique for each user)
Location (contains city, state and country separated by commas)
Age

- Shape of Dataset - (278858, 3)

❖ Books_dataset.

ISBN (unique for each book)
Book-Title
Book-Author
Year-Of-Publication
Publisher
Image-URL-S
Image-URL-M
Image-URL-L

- Shape of Dataset - (271360, 8)

❖ Ratings_dataset.

User-ID
ISBN
Book-Rating

- Shape of Dataset - (1149780, 3)

Proposed Work :

Our main aim is to provide the best book recommendations to our users. We use a hybrid recommender system for this purpose. Hybrid systems are the combination of two other types of recommender systems: content-based filtering and collaborative filtering. Using collaborative filtering we recommend some books to the user on login. Using content based filtering we recommend books based on the book name searched by the user. We use the K-means clustering algorithm to cluster books based on rating. To extract the relevant text to include in the audio book, we use Jenks Optimization Algorithm and we generate the book with the help of Google Vision API.

Exploratory data analysis (EDA):

- Visualization on some important parts like most rated books, most popular books, most popular authors, most number of ratings

Data Preprocessing:

- Data cleansing
- Nan value treatment
- Extraction of relevant featuress

Model building:

- Splitting data into train and test set
- Neural Network creation(Using functional API)
- Training on train data and validating on test data

Recommendation:

- Loading model
- Creating a function for recommendation

Recommendation System :

In this section we have described how each of the algorithms/methods mentioned in the previous section have been utilized effectively with respect to our project.We use a hybrid recommender system which is the combination of two other types of recommender systems: content-based filtering and collaborative filtering.

Collaborative filtering:

Collaborative filtering is a method by which user ratings are used in order to determine user or item similarities[2].For collaborative filtering, the primary features necessary are user_id, book_id, and ratings. Collaborative filtering does not need the features of the items to be given. Every user and item is described by a feature vector. It collects user feedback on different items and uses them for recommendations [2].We have applied K-means clustering and Gaussian mixture modelling to cluster the users and reach the best silhouette score. We have used the euclidean similarity distance measure. The formula for Euclidean similarity distance measure is given as follows:

$$\frac{1}{1 + d(p_1, p_2)}$$

Content-Based filtering :

Content-based filtering is a method of recommending items by the similarity of the said items..A content-based recommender system aims to predict a user's characteristics or behaviour based

on the item's characteristics, which he or she responds positively to[2]. For example, if a user likes the first book of Harry Potter, and if the second book is similar to the first, then it recommends the second book to the user.We use cosine similarity and our item similarities are a combination of user ratings and features derived from books themselves. The formula for cosine similarity is given as:

$$\cos(x, y) = \frac{(x \bullet y)}{\|x\|\|y\|}$$

Memory based based Collaborative filtering :

e.g. KNN (K-Nearest Neighbors)

Utilizes entire user-item rating information to calculate similarity scores between users or items.

In user based collaborative filtering, two user's are considered similar, if they rate items in a similar manner. An item is recommended to a user, if another user i.e., similar to the user in question has liked the item.

In item based collaborative filtering, two item's are considered similar, if users rate them in a similar manner. An item is recommended to a user, that is similar to the items the user has rated in the past.

Model based based Collaborative filtering :

e.g. Singular Value Decomposition

Model based approach utilizes information from the dataset to build a model & the model (not the entire dataset) is thereafter used for making recommendations .

For dimensionality reduction, the model optimizes certain latent features (hidden characteristics) that are able to map user item preferences In the process, the model is able to predict an estimated rating for all user-item pair, where user has not yet rated an item.

Conclusion:

All of our systems– purely content-based, purely collaborative-filtering, and hybrid– performed quite well. Looking back on the project, one thing that we might have chosen to do differently in retrospect would have been to spend more time searching for a dataset of ratings with a higher rating variance per user. Had we been able to find such a dataset, our implementations of algorithms would have been tested on data that would have been more representative of what a typical commercial recommendation system could access in creating its predictions. However, given the data that was available to us, as well as the results our various approaches produced, our systems were largely successful, providing insight into how the different systems we regularly use work and the varying algorithms that make that possible.

This project was able to present a comprehensive review on research previously targeted on improving recommender systems. From the results and visualizations, we can deduce that the accuracy of rating followed a normal distribution which suggests consistency and efficiency.

We were able to train efficient models that had high accuracy. The model performed well on random inputs outside

the dataset as well, we had tried to generate an audiobook for a research paper. We also tried to generate an audiobook, given a handwritten document. The former and later experiments gave us accuracy of above 70%.

The System has adequate scope for improvement in the future. We can develop and launch a Mobile app. Also System security, data security and reliability could be the major features which can be done in future.We can also add the API for the shopping and payment gateway. In the existing system there are only some selected categories like we search using book names, so as an extension to the site we canadd more categories. Also we can add admin side with some functionalities like books management, User management etc.

References:

Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. Springer, 2011.

Jessica E. Moyer, “Audiobooks and E-books: A Literature Review'',2012 Reference & User Services Quarterly, vol. 51, no. 4, pp. 340–54.

Badrul Sarwar et al. “Item-based collaborative filtering recommendation algorithms.” In: Proceedings of the 10th international conference on World Wide Web. ACM. 2001, pp. 285–295.

K. Tsuji, F. Yoshikane, S. Sato and H. Itsumura, "Book Recommendation Using Machine Learning Methods Based on Library Loan Records and Bibliographic Information," 2014 IIAI 3rd International Conference on Advanced Applied Informatics, pp. 76- 79.

Raymond J Mooney and Loriene Roy. “Content-based book recommending using learning for text categorization.” In: Proceedings of the fifth ACM conference on Digital libraries. ACM. 2000, pp. 195–204.