Project Title: Personalised Book Recommendation System

# Team:

1. Amul Naik: 12410068
2. Divya Agarwal 12410061
3. Medha Adhikari: 12410093
4. Satyajit Chakraborty: 12410076

# Project Statement:

We aim to develop a sophisticated book recommendation system using collaborative filtering techniques, specifically user-based and item-based collaborative filtering. This system will recommend books to users based on their past ratings and the ratings of others. We'll enhance the recommendation process by employing similarity measures such as Cosine Similarity and Pearson Correlation to assess the similarity between books or users based on their ratings.

Furthermore, we'll use PCA (Principal Component Analysis) for dimensionality reduction and t-SNE

(t-distributed Stochastic Neighbour Embedding) for visualizing clusters of books and users in a lower- dimensional space. These techniques will help us better understand the underlying structure of the data, improving the accuracy and interpretability of our recommendations.

# Dataset Source and Description:

We will utilize the Books Recommendations Dataset from Kaggle: <https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset>

The dataset includes the following files:

1. books.csv: Contains metadata for books, including book IDs, titles, and descriptions.
2. users.csv: Includes user data such as user IDs and their details (e.g., location).
3. ratings.csv: Contains ratings provided by users for various books, linking each user ID, book ID, and the rating.

The dataset comprises: 278,858 users who have provided 1,149,780 ratings (explicit / implicit) about 271,379 books.

This dataset is ideal for building a collaborative filtering-based recommendation system, given the extensive ratings from multiple users for numerous books, allowing us to compute similarity metrics and make recommendations.

# Research Question:

The key question for this project is: "*How can we develop an effective book recommendation system using collaborative filtering to predict books that users will enjoy based on their past ratings and similarities with other users or books?*"

By leveraging Cosine Similarity and Pearson Correlation, we'll calculate similarities between users or books. Additionally, we'll use PCA and t-SNE to reduce the data's complexity and help identify patterns within the data, aiding in better understanding user and book clusters.

# Implications of Answering the Research Question: Addressing this question and creating a book recommendation system entails several benefits:

1. **Personalization:** Enhancing user satisfaction by providing personalized book suggestions that align with individual preferences.
2. **Improved Discovery of New Books**: Helping users discover books that might otherwise be overlooked, while aiding businesses in highlighting books of interest to specific user segments.
3. **Enhanced User Engagement:** Increasing engagement on the platform by recommending books users are likely to enjoy, promoting more book purchases, readings, or rentals.
4. **Business Insights**: Gaining valuable insights into user behaviour and preferences, aiding in shaping marketing strategies and content curation.
5. **Competitive Advantage and Business Growth:** Improving sales and revenues through higher engagement and purchases on the platform as relevant books are recommended to users, especially relevant in a crowded market.