****

**Business Objective:** Generate the features from the dataset and use them to recommend the booksaccordingly to the users.

Collaborators: Syed Farhanuddin, Arindam Bhramachary

**Introduction to Recommendation System - Books**

Recommender Systems are algorithms that aims at suggesting relevant item to users, for example, movies, books, products, etc. It is critical in some industries for it can generate profits in huge volume when efficient. It can help in standing out significantly from competitors.

In this project, we are building a Recommendation system for Books by leveraging Popularity Based and Collaborative Filtering models to three datasets namely, Books.csv, Users.csv & Ratings.csv

**Exploratory Data Analysis (EDA)**

Analysis on some important parts from dataset like most rated books, most popular books, most popular authors, high and low ratings, number of ratings, etc.

* Describing the dataset
* Cleaning the data
* Checking invalid records
* Missing value detection and imputation
* Duplicated records
* Outliers and Correlation

**Data Visualization**

Data Visualization, the graphical or visual representation of data. The idea behind data visualization is to make it easier for us (humans) to identify trends, outliers, and patterns in data. Plots used for Data Visualization:

* Scatter plot
* Bar graph
* Line graph
* Pie chart
* Histograms
* Heatmap
* Box-Plot

**Feature Engineering**

Implemented this technique to transform data into more effective set of inputs, that act as catalysts for recommendation. The methods opted

* Feature Creation: Construction of new features based on existing as needed, as needed.
* Feature Selection: Selection of relevant features for recommendation.
* Feature Transformation: Converting features into suitable format per requirement for the model, as needed.

**Model Building - Recommendation System**

A model captures underlying patterns and relationships of data, apart from simply memorizing training data. Once built, the model can be applied to new data to make predictions and obtain results. The recommendation models built.

Popularity Based: This system suggests items that are currently popular or trending to users.

Collaborative Filtering: A personalized recommendation strategy that uses data from other users to identify similarities between users and provide recommendations.

**Hyperparameter Tuning**

These hyperparameters control various aspects of the learning process, such as

* the learning rate or
* the number of neurons in neural network or
* the kernel size in the support vector machine etc.

Hyperparameter tuning can be performed through GridSearchCV, KNeighbors Classifier or Nearest Neighbors when the difference in factors such as accuracy or RMSE is minimal to zero. Thus, hyperparameter tuning is performed with limited requirements as the model/s operate at the desired rate.

**Model Evaluation**

For regression evaluation,computedR-Squared, Mean Squared Error and Root Mean Squared Error.

For classification evaluation,computedAccuracy, Precision, Confusion Matrix and Classification Report.

**Model Deployment**

This isa process of making machine learning models available to users, developers, and systems. It allows users to make business decisions based on data, and applications to interact with the model.

In Python, libraries like Pickle or job lib can serialize the model and save it to a file. This file can then be loaded into a Flask application to make predictions.

Pickle: This is primarily used in serializing and deserializing a Python object structure or a process of converting a Python object into a byte stream to store it in a file/database, maintain program state across sessions, or transport data over the network.

Flask: This is a web framework that allows developers to build lightweight web applications quickly and easily with Flask Libraries.