**PROJECT 1:**

**TASK SELECTED:**

1. Build a recommender system that recommends books to read for every user based on their personal tastes and previous book ratings.

**DATA PREPARATION:**

**CODE SCREEN SHOTS:**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated**

**CODE:**

import pandas as pd

ratings\_df = pd.read\_csv('Ratings.csv', sep=';', dtype={'User-ID': int, 'ISBN': str, 'Rating': int})

print(ratings\_df)

from scipy.sparse import coo\_matrix

user\_mapping = {user\_id: idx for idx, user\_id in enumerate(ratings\_df['User-ID'].unique())}

book\_mapping = {isbn: idx for idx, isbn in enumerate(ratings\_df['ISBN'].unique())}

ratings\_df['User-ID-Mapped'] = ratings\_df['User-ID'].map(user\_mapping)

ratings\_df['ISBN-Mapped'] = ratings\_df['ISBN'].map(book\_mapping)

user\_ids = ratings\_df['User-ID-Mapped']

book\_ids = ratings\_df['ISBN-Mapped']

ratings = ratings\_df['Rating']

sparse\_matrix = coo\_matrix((ratings, (user\_ids, book\_ids)))

from sklearn.datasets import dump\_svmlight\_file

dummy\_labels = [0] \* sparse\_matrix.shape[0]

temp\_file\_path = 'temp\_user\_book\_matrix.libsvm'

dump\_svmlight\_file(sparse\_matrix, dummy\_labels, temp\_file\_path, zero\_based=True)

output\_file\_path = 'user\_book\_matrix.libsvm'

with open(temp\_file\_path, 'r') as infile, open(output\_file\_path, 'w') as outfile:

for line in infile:

stripped\_line = ' '.join(line.split()[1:])

if stripped\_line.strip():

outfile.write(stripped\_line + '\n')

print(f"Sparse matrix saved without labels in LibSVM format: {output\_file\_path}")

for i in range (0,len(user\_ids)):

print( ratings\_df['User-ID'][i],user\_ids[i])

for i in range (0,len(book\_ids)):

print( ratings\_df['ISBN'][i],book\_ids[i])

**STEPS IMPLEMENTED:**

* First, we read the ratings csv file with ‘;’ as the delimiter into a dataframe object
* Then we create a dictionary user\_mapping and book\_mapping that will map the user\_ids and isbns to new ids starting from 0 for the first id.
* We then form the new ids formed to create the user\_ids and book\_ids variables which are the values of the user\_mapping and book\_mapping dictionaries formed earlier. Which is nothing but assigning a number to each category so they can be represented in the form of a libsvm file
* The categories are then converted into a spare matrix using the coo\_matrix object. Doing so will result in a sparse matrix like the one given in the Assignment page. <https://canvas.asu.edu/courses/197512/assignments/5710096>
* The spare matrix is then converted to a libsvm file using the dump\_svmlight\_file function of sklearn.
* The format of the converted libsvm file is not of the desired format so we are reading the libsvm format and removing the unnecessary labels in the libsvm file to convert the file to the desired format
* The below are the examples format of how the libsvm file looks
* We can also read and alter the sparse matrix to remove the labels before saving as a libsvm file but that takes a lot of time and is much slower than the currently used approach.
* Then we print and see the original user ids and their corresponding new user ids created.
* We also print the old book ids and their corresponding new book ids created.

A screenshot of a computer

Description automatically generated

**FORMAT OF TEMP FILE:**

A screenshot of a computer

Description automatically generated

**FORMAT OF THE FINAL FILE:**

A screenshot of a computer

Description automatically generated