**“IFT 511 – Analyzing Big Data**

**TEAM - 05**

**Project Step 1: Task Selection and Data Cleaning**

Narasimha Uday Kiran Narayana- 1232310107

Sruthi Bijja – 1232454316

Mahan santosh satya sai ashish bandaru- 1231818434

Rohith Santosh Reddy Dasari - 1232222656

**Asmaa Elbadrawy**

**November 17, 2024”**

**Topic:**

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

**Code:**

import pandas as pd

from collections import defaultdict

# File paths

ratings\_file\_path = r"C:\Users\nnuda\Downloads\archive (3)\Ratings.csv"

users\_file\_path = r"C:\Users\nnuda\Downloads\archive (3)\Users.csv"

# Load Ratings dataset

ratings\_df = pd.read\_csv(ratings\_file\_path, sep=';', names=['User-ID', 'ISBN', 'Rating'], skiprows=1)

# Load Users dataset with explicit dtype specification

users\_df = pd.read\_csv(

users\_file\_path,

sep=';',

names=['User-ID', 'Location', 'Age'],

skiprows=1,

dtype={'User-ID': str, 'Location': str, 'Age': 'float32'} # Explicit types

)

# Mapping each unique User-ID and ISBN starting from 1

user\_to\_index = {user: i + 1 for i, user in enumerate(ratings\_df['User-ID'].unique())}

isbn\_to\_index = {isbn: i + 1 for i, isbn in enumerate(ratings\_df['ISBN'].unique())}

# Dictionary to store each user's ratings, {isbn\_index: rating}

user\_ratings = defaultdict(dict)

# Populate the user\_ratings dictionary with sequential user and ISBN indices

for \_, row in ratings\_df.iterrows():

user\_idx = user\_to\_index[row['User-ID']]

isbn\_idx = isbn\_to\_index[row['ISBN']]

user\_ratings[user\_idx][isbn\_idx] = row['Rating']

# Write the LIBSVM formatted output to a file

output\_file\_path = r"C:\Users\nnuda\Downloads\user\_booklibsvmnew.libsvm"

with open(output\_file\_path, 'w') as f:

for user\_idx in sorted(user\_ratings):

ratings = [f'{isbn\_idx}:{rating}' for isbn\_idx, rating in sorted(user\_ratings[user\_idx].items())]

line = ' '.join(ratings) + '\n'

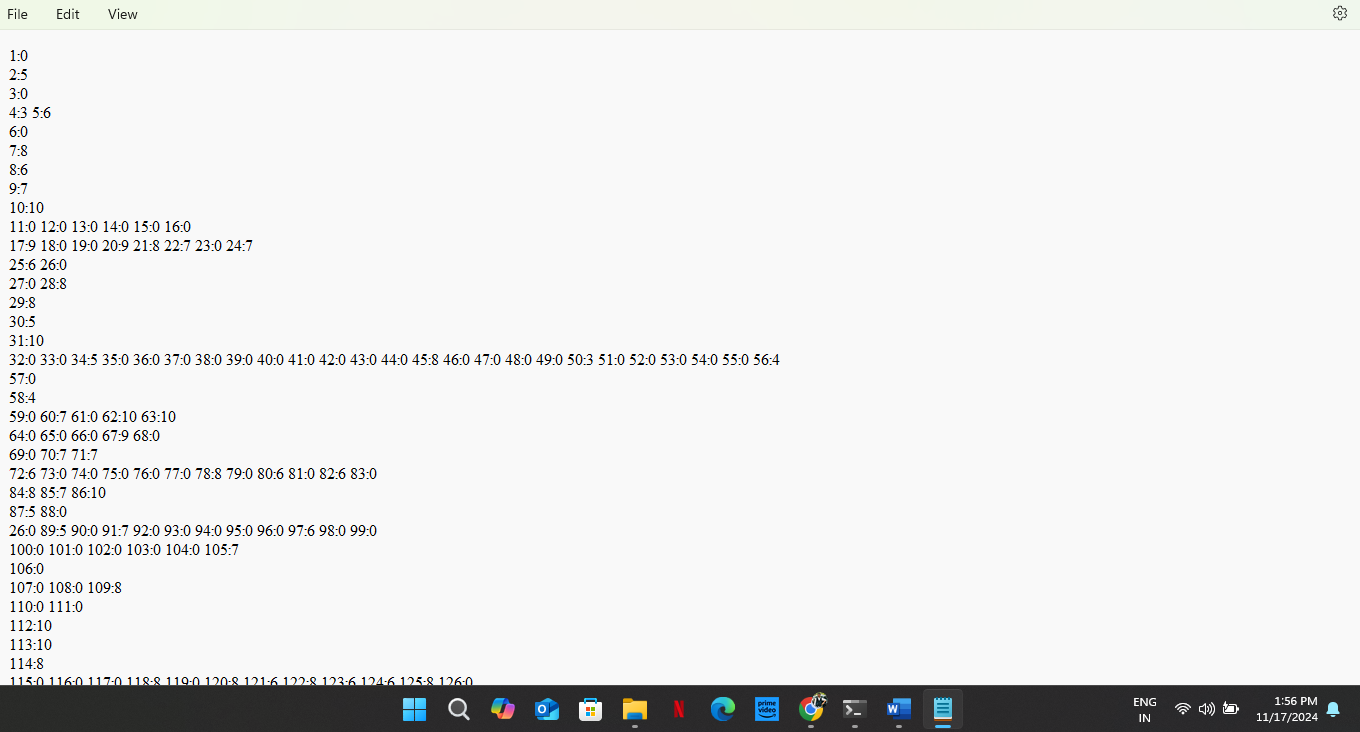
f.write(line)

print(f"File '{output\_file\_path}' written successfully!")

A screenshot of a computer

Description automatically generated

**LIBSVM Screenshot:**

****