Course Title: **IFT 511 – Analyzing Big Data**

**Team Number:** 29

**Team Members:**

Archana Kanchimireddy-1232309145

Yasaswi Pandaraboyina - 1231868705

Venkata Naga Sai Chakradhar Kollipara – 1232332129

Raghu Menni Lokanadhanaidu-1232328671

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

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

**Due Date:** November 10, 2024

**Task 1**:

**Code**:

import pandas as pd

from collections import defaultdict

# Loading the dataset

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

# 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, without user ID

with open('user\_booklibsvmnew.libsvm', 'w') as f:

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

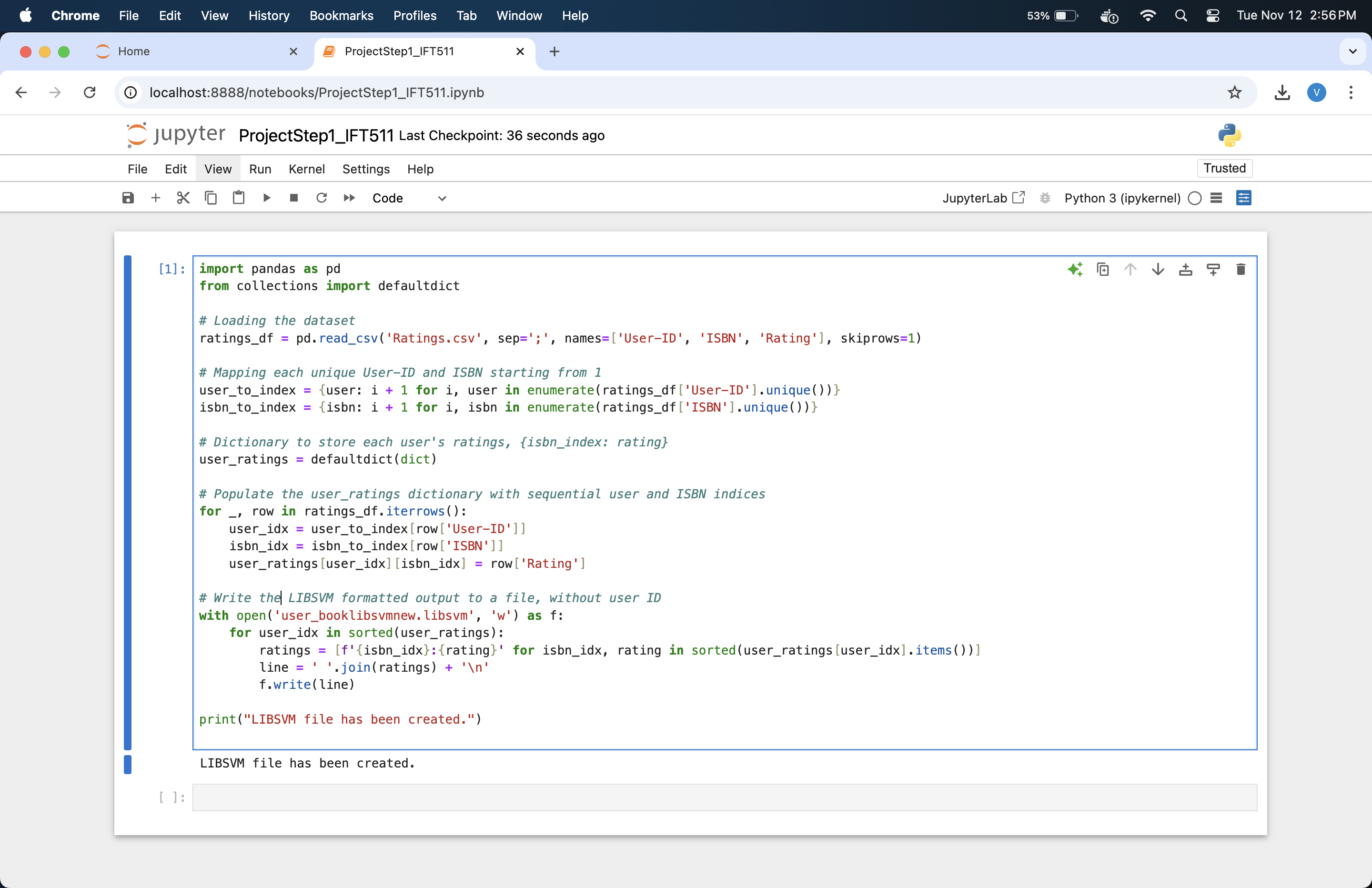
ratings = [f'{isbn\_idx}:{rating}' for isbn\_idx, rating in sorted(user\_ratings[user\_idx].items())]

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

f.write(line)

print("LIBSVM file has been created.")

**Screenshot of code**:



**LibSVM file Screenshot**:

