CSCI316 Big Data Mining Implementation and Techniques

Laboratory 1

Objective

- To Explore Google Colab
- To practice basic Python programming for data analytics

Introduction

Google Colab (Colaboratory) is a free cloud-based Jupyter notebook environment that runs entirely in your browser. It requires no setup to use and offers free access to computing resources, including GPUs and TPUs, making it ideal for machine learning and data science tasks.

Step 1: Accessing Google Colab

1. **Open Your Web Browser:** Use Google Chrome for the best compatibility and performance, though other modern browsers will generally work.

2. **Go to Google Colab:** Navigate to colab.research.google.com.

3. **Sign In:** If you're not already signed in to your Google account, you'll be prompted to do so. Colab is tightly integrated with Google Drive, where your notebooks will be saved.

Step 2: Creating and Opening Notebooks

1. **Create a New Notebook:**

   - Once on the Colab homepage, you'll see a pop-up. Select "New notebook" from this dialog.

   - Alternatively, go to File > New notebook in the menu bar.

   - A new, empty notebook will open, ready for your code.

2. **Open an Existing Notebook:**

   - From the initial pop-up, you can select "Recent" to open notebooks you've worked on recently.

   - Choose "Google Drive" to browse your Google Drive for .ipynb files.

   - Select "GitHub" to open notebooks directly from a GitHub repository (useful for course materials).

   - Select "Upload" to upload a .ipynb file from your local computer.

Step 3: Understanding the Notebook Interface

A Colab notebook consists of "cells." There are two main types:

1. **Code Cells:** Where you write and execute Python code.

   o **Execution:** Click the "play" button (triangle icon) on the left of the cell, or press Shift + Enter.

   o **Output:** The output of your code (e.g., print statements, plots) will appear directly below the cell.

2. **Text Cells (Markdown Cells):** Where you write explanations, notes, headings, and format text using Markdown.

   o **Editing:** Double-click a text cell to enter editing mode.

   o **Rendering:** Press Shift + Enter or click outside the cell to render the Markdown.

   o **Markdown Basics:** Learn basic Markdown for headings (#, ##), bold (**text**), italics (*text*), lists (- item), etc.

Toolbar and Menus:

- **File:** Save, Open, New notebook, Download, Print.

- **Edit:** Cut, Copy, Paste cells, Find and replace.

- **View:** Toggle table of contents, code line numbers.

- **Insert:** Add code cell, Add text cell, Add section header.

- **Runtime:** The most important menu for labs.

   o **Run all:** Executes all cells in the notebook.

   o **Run selected cells:** Executes highlighted cells.

   o **Change runtime type:** Allows you to select hardware accelerators (None, GPU, TPU). Crucial for deep learning labs.

   o **Restart runtime:** Clears all variables and resets the environment.

   o **Disconnect and delete runtime:** Releases computing resources.

- **Table of Contents (Left Pane):** Automatically generated from Markdown headings, useful for navigation.

- **Files (Left Pane):** Access your files, upload/download, mount Google Drive.

Step 4: Running Code and Using Resources

1. **Basic Code Execution:**

2. print("Hello, Colab!")

3. a = 5

4. b = 10

5. print(a + b)

Type this into a code cell and run it.

6. **Installing Libraries:** You can install any Python library using pip within a code cell:

7. !pip install pandas matplotlib

The ! tells Colab to run the command as a shell command.

8. **Importing Libraries:**

9. import pandas as pd

10. import numpy as np

11. **Connecting to Google Drive:** This is essential for accessing your datasets or saving large outputs.

12. from google.colab import drive

13. drive.mount('/content/drive')

This will prompt you to authorize Colab to access your Google Drive. Once mounted, your Drive contents will be accessible at /content/drive/MyDrive/.

14. **Using Hardware Accelerators (GPU/TPU):**

    o Go to Runtime > Change runtime type.

    o Under "Hardware accelerator," select "GPU" or "TPU" if your lab requires it.

    o Click "Save."

    o You can verify if a GPU is active by running:

    o !nvidia-smi

(This command works for GPU, not TPU).

Step 5: Saving and Sharing Your Work

1. **Saving:**

    o Colab automatically saves your notebook to your Google Drive (MyDrive/Colab Notebooks/ by default).

    o You can manually save via File > Save or File > Save a copy in Drive.

2. **Downloading:**

   o File > Download > Download .ipynb: Saves the notebook file to your computer.

   o File > Download > Download .py: Saves only the code cells as a Python script.

3. **Sharing:**

   o Click the "Share" button (top right corner).

   o You can share it like any other Google Drive document (via link or specific email addresses) with various permissions (viewer, commenter, editor).

   o Remember that collaborators can run your code and see outputs.

Step 6: Best Practices for Labs

- **Organize with Text Cells:** Use Markdown cells to clearly label sections (e.g., "Part 1: Data Loading," "Part 2: Model Training," "Results").

- **Add Comments:** Comment your code (# This is a comment) to explain logic.

- **Run Cells Sequentially:** Ensure cells are run in the correct order, especially if one cell's output is an input for another.

- **Save Regularly:** Although autosave is present, manually saving before closing or major changes is good practice.

- **Restart Runtime if Issues Arise:** If you encounter unexpected errors or memory issues, try Runtime > Restart runtime to get a clean slate.

- **Monitor Resources:** Be mindful of resource limits, especially for free GPU/TPU usage. If you get disconnected, it might be due to inactivity or hitting a usage cap.

Group Formation Guidelines for Assignment Task

1. **Group Size**

   o Each group must have **3 to 5 members**.

   o **Not smaller than 3**, and **not larger than 5**.

2. **Student Type (Full-time / Part-time)**

   o All members in a group must be either:

      ▪ **All full-time students**, or

      ▪ **All part-time students**.

- o Mixed groups (full-time + part-time) are **not allowed**.

3. **Cross-Tutorial Group Formation (For Full-time Students Only)**

    - o Full-time students **may form groups with students from other tutorial groups**.

    - o However, the **presentation schedule** will be based on the tutorial group that has the **most members** in your group.

        - ▪ Example: If 3 out of 4 members are from Tutorial Group A, your presentation will follow Group A's schedule.

    - o **All group members must attend and present** at the assigned time.

4. **Important Notes**

    - o Ensure your group meets the size and student-type requirements.

    - o Plan presentation availability in advance if forming cross-tutorial groups.

    - o Failure to comply with these rules may affect your assignment submission.

Next Steps:

- • Discuss and finalize your group members early.

- • Confirm group details with your tutor if needed.