**CS 123A Project README**

Project Title: Small Molecule Drug Development

This README document contains 3 main sections:

* Project Submission Folder Structure
* Instructions to Run Preprocessing and Modeling Code
* Instructions to Run BRAF Inhibition Prediction Interface Code

# Project Submission Folder Structure

Upon opening the submitted ZIP file, there should be the following:

* “Code” folder
* Presentation Slides
* Presentation Slides in PDF format (4-pages)
* Hands-on Activity
* Project Report
* chemical\_compounds.csv
  + CSV file containing the chemical compound dataset we used to perform EDA, preprocessing, and modeling on

The folder "Code" contains the following:

* Chemical\_Compounds\_EDA\_Findings.ipynb
  + Python notebook containing code that perform Exploratory Data Analysis (EDA)
  + This Python notebook is not meant to be ran, but viewed as a notebook detailing our observations
* CS\_123\_Preprocessing\_Modeling
  + Python notebook that performs Preprocessing and Modeling
* Folder named “UI”
  + Folder containing the code and or format for the UI interface
  + Instructions to run the interface will be detailed in a section below

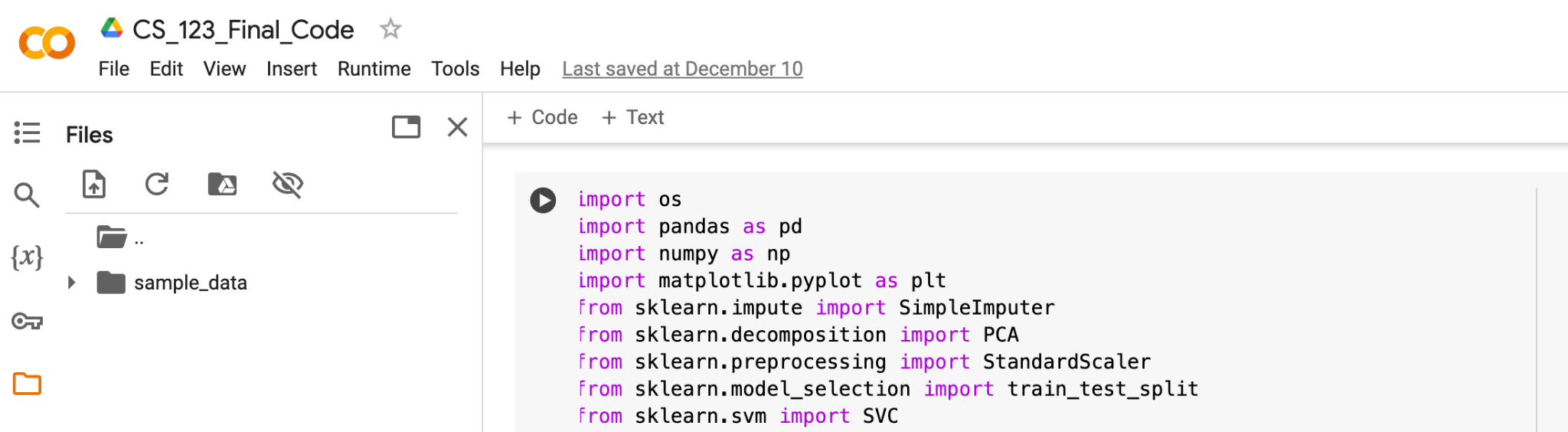
The folder “UI” contains the following:

* “templates” folder
  + Folder that contains HTML pages that will be displayed for the interface
* “SampleInput” folder
  + Folder that contains the Sample Input a user can put into the interface, more specifically the “cc1\_test1\_X.csv”
  + “cc1\_test1\_y.csv” is not meant to be used as an input but as a reference what the actual label is for features depicted in “cc1\_test1\_X.csv”
  + In the case where you want to use your own input, ensure the chemical compound input has all the same columns as “chemical\_compounds.csv” has but without the column “Class.” Your input will need to be saved as a CSV file and contain the column headers.
* “PreprocessingObjects” folder
  + Folder that contains preprocessing objects to preprocess the user’s input and set up so the models can perform prediction on it
* “Models” folder
  + Folder that contains SVM models
* BRAF\_Inhibitor\_Interface.ipynb
  + Code that will run the interface for take in user’s input and give prediction(s)

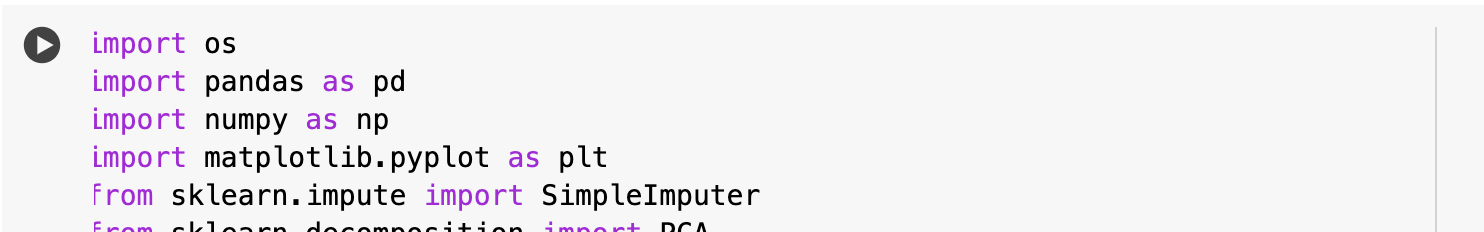
# Instructions to Run the Preprocessing and Modeling Code:

To run the preprocessing and modeling code, do the following steps:

1. Import or upload the “CS\_123\_Preprocessing\_Modeling” code into Google Drive or Google Colab
2. Upon opening the code in Google Colab, click the folder icon located on the left sidebar, so that it expands and displays as shown below:



1. Download the chemical\_compounds.csv
2. Click the Upload icon and upload the chemical\_compounds.csv
3. Once the upload has been completed, Click on the play icon in the code block to execute the preprocessing and modeling based on the data in chemical\_compounds.csv

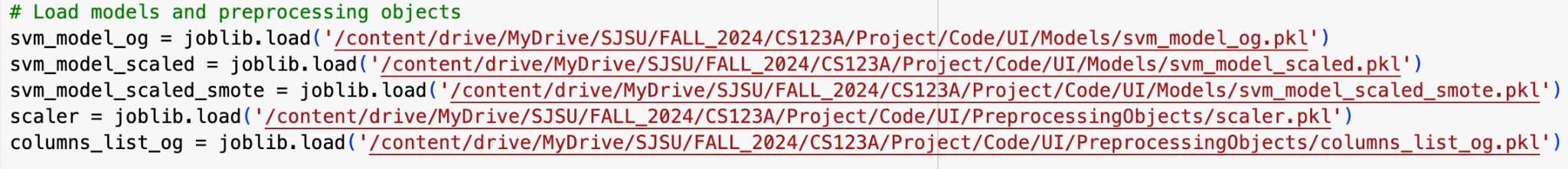


1. The output will be displayed directly below the code block after executing the code block

# Instructions to Run the BRAF Inhibition Prediction Interface Code:

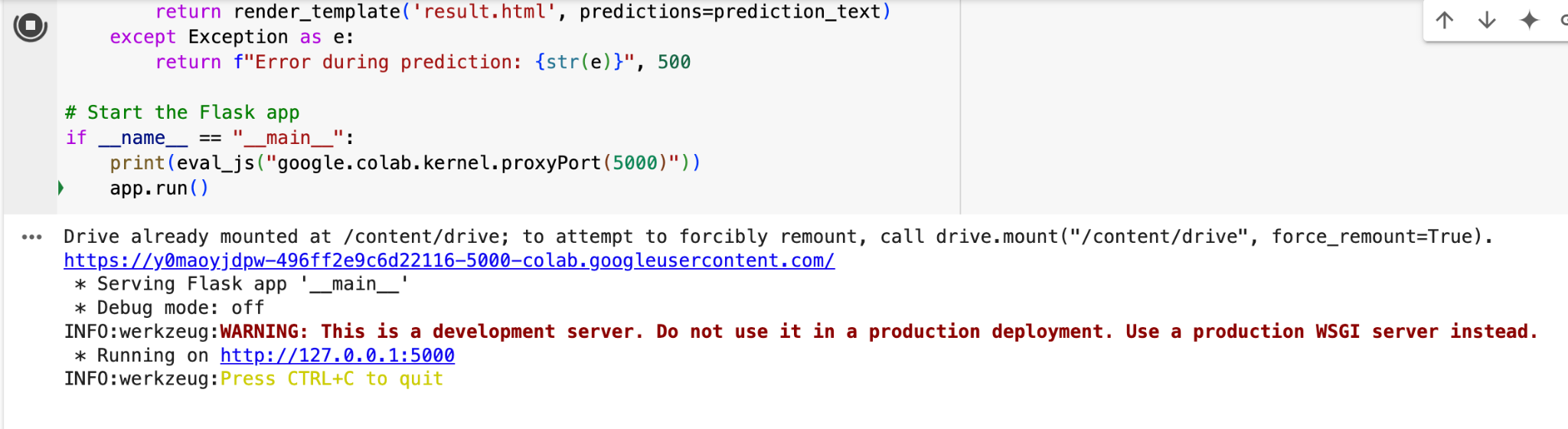
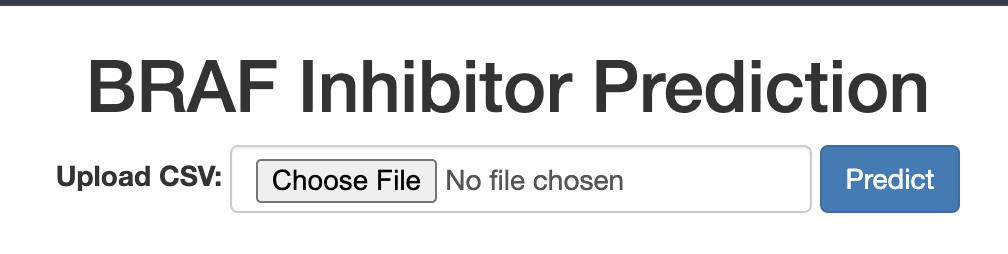
**To run the interface, do the following steps:**

1. Import the “UI” folder into your Google Drive
2. Open BRAF\_Inhibitor\_Interface.ipynb
3. Update the drive path to the Model objects and Preprocessing Objects. The models can be found in the “Models” folder while the scaler and column\_list\_og can be found in the “PreprocessingObjects” folder:



1. Update the drive path to the “templates” folder. The folder can be found in the UI folder.



1. Run this code block only
2. While this is running, look below the code block and look for a link:  
   
3. Click on the first link and it should open a new tab displaying the interface as shown below:  
   
4. Use the cc\_test1\_x.csv as a sample input. This can be found in the “SampleInput” folder.
5. After uploading and submitting the input file, the output would display our models’ prediction, where Model 1 is the model with no scaling and Model 2 has scaling:

