National College of Ireland

Configuration Manual

MSc Research Project
Programme Name

Sumit Jeevan Mandal
Student ID: 24138371

School of Computing
National College of Ireland

Supervisor: Victor del Rosal

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Sumit Jeevan Mandal
Student ID:	24138371
Programme:	MSc in Artificial Intelligence for Business **Year:** 2025
Module:	Practicum
Lecturer:	Victor del Rosal
Submission Due Date:	28-01-2026
Project Title:	Enhancing Lead Scoring and Prioritization in Ed-Tech companies Using Supervised Machine Learning
Word Count:	773 **Page Count:**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Sumit Jeevan Mandal
Date:	28-01-2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Sumit Jeevan Mandal
Student ID: 24138371

1 Introduction

This configuration manual aims to act as a detailed guide for the Lead Scoring Automation project. It not only provides the means for setting up the development environment but also for installing necessary packages and running the source code as well. This document serves as a safeguard for the reproducibility of the research findings as well as the possibility of deploying the artefact on any standard machine with the help of the Google Colab platform.

2 System Requirements

Project implementation through Google Colab enabled the utilization of the cloud, hence the local system requirements are quite low.

2.1 Hardware Requirements

- Device: Laptop or Desktop Computer
- RAM: At least 4GB (8GB recommended for smooth operations of the browser)
- Processor: Intel Core i3 or similar and above
- Internet Connection: A reliable high-speed internet connection is mandatory for connecting to the Google Colab runtime.

2.2 Software Requirements

- Operating System: Windows 10/11, macOS, or Linux
- Web Browser: Google Chrome (Recommended), Mozilla Firefox, or Microsoft Edge
- Google Account: Access to Google Drive and Colab requires a valid Google account.

3 Environmental Setup

The project is based on Google Colab, which is a hosted Jupyter notebook service that does not need any setup to work. Google controls the runtime environment.

- Platform: Google Colab
- Programming Language: Python 3.10+
- Runtime Type: Standard (CPU is adequate, GPU is not needed for this dataset size).

3.2 Required Libraries

There are several Python libraries used in the project which are listed below. The code has automated commands to install missing libraries but for your reference, they are included here:

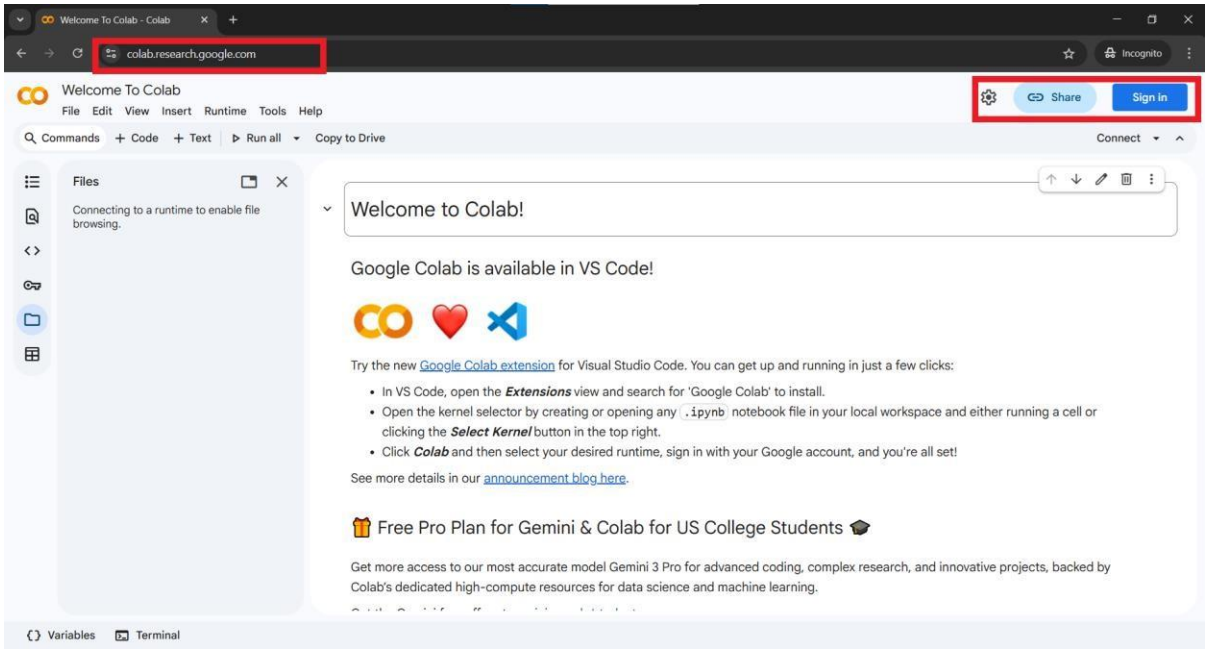
- pandas (Managing data)
- numpy (Math operations)
- matplotlib & seaborn (Graphical representation of data)
- scikit-learn (ML models and steps)
- xgboost (Boosting method)
- lightgbm (Light gradient boosting framework)
- imbalanced-learn (SMOTE for class imbalance)

4 Environmental Setup

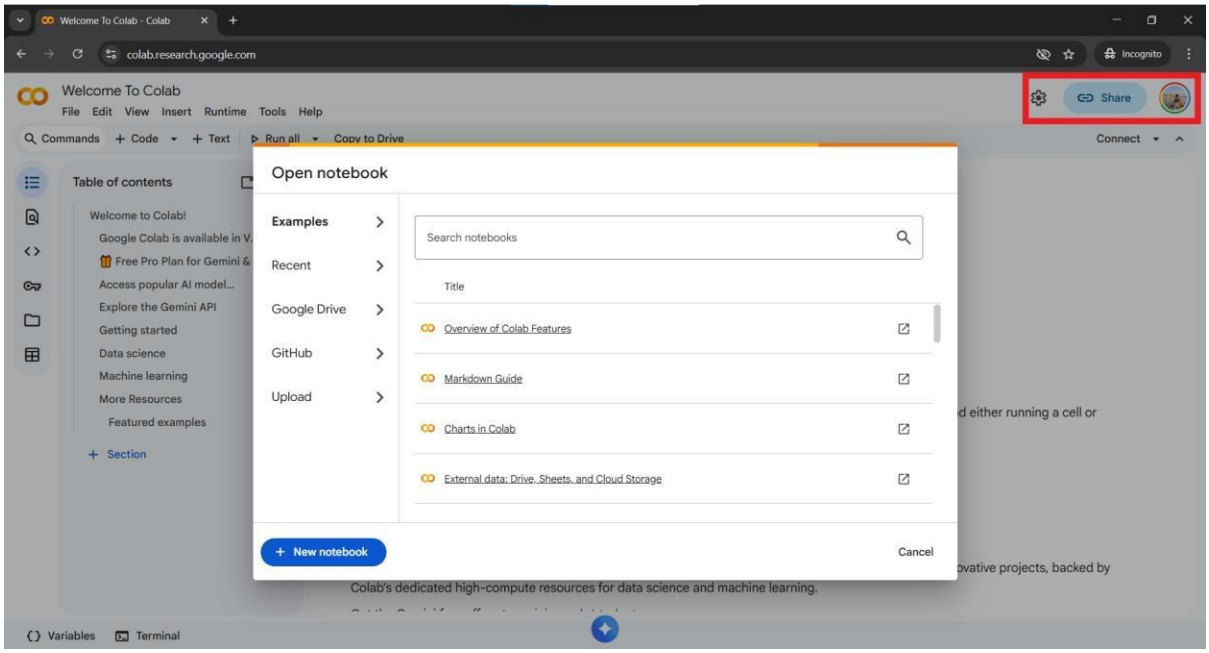
Step 1: Access Google Colab

1. Open your web browser.

2. Go to https://colab.research.google.com/.

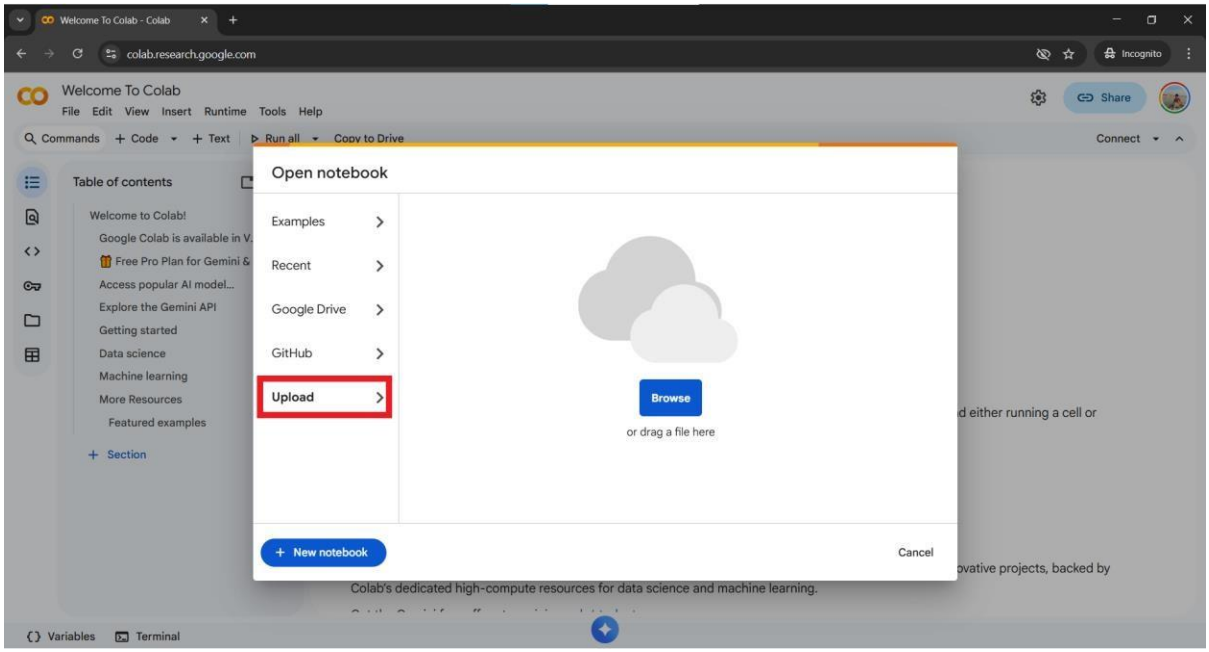


3. Log in with your Google Account.

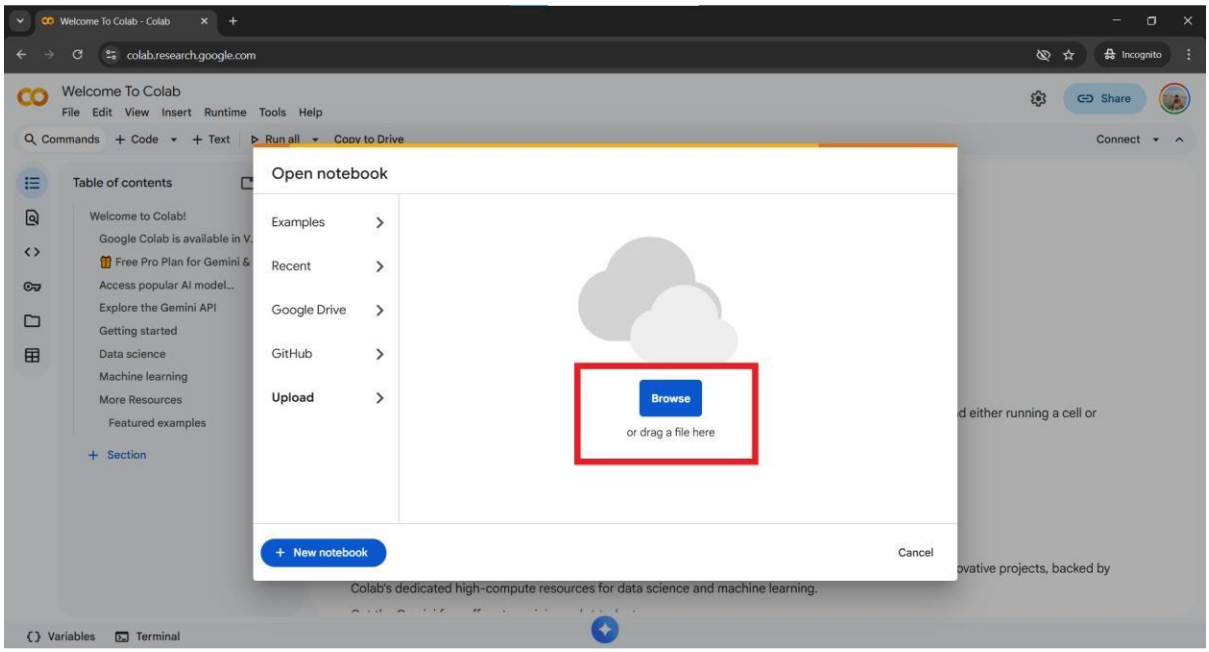


Step 2: Upload the Notebook

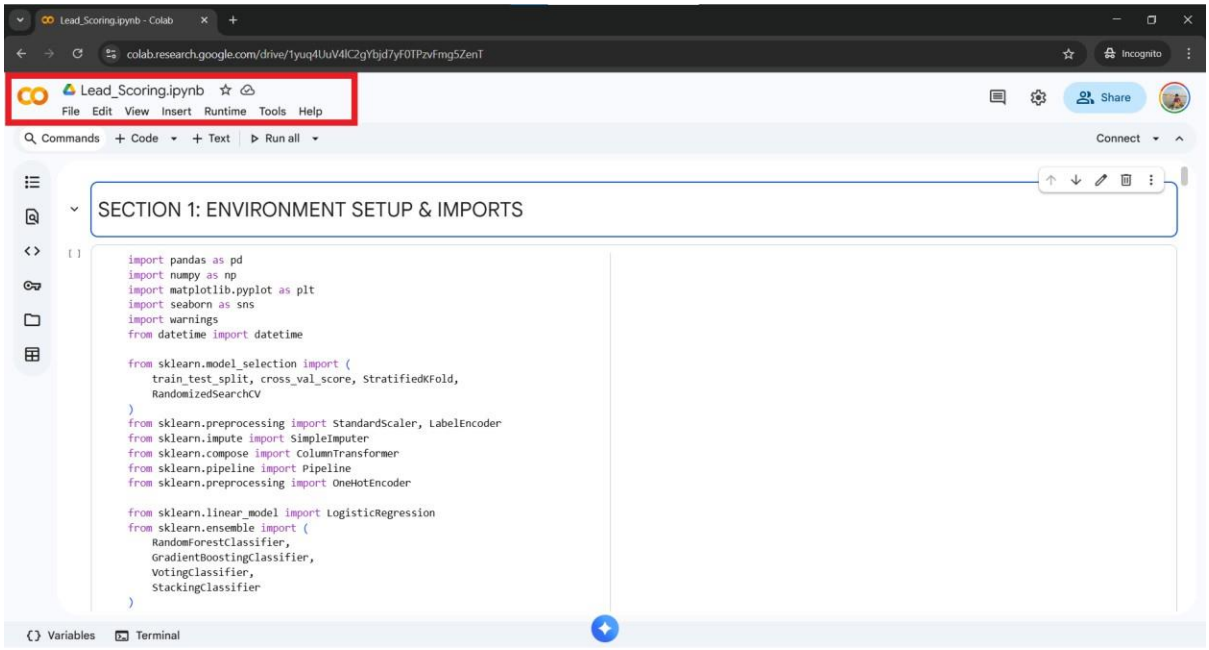
1. In the window called "Open notebook", click on the Upload tab.



2. Click Browse and select the file Lead_Scoring.ipynb.



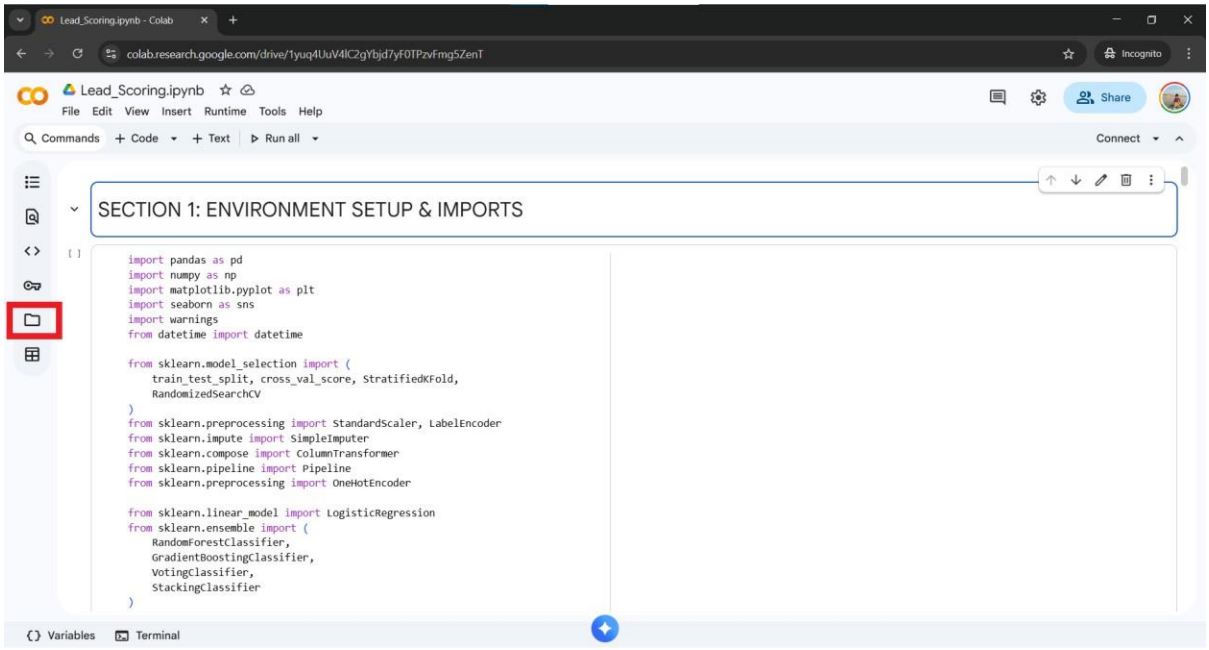
3. The notebook will be displayed in the Colab environment.



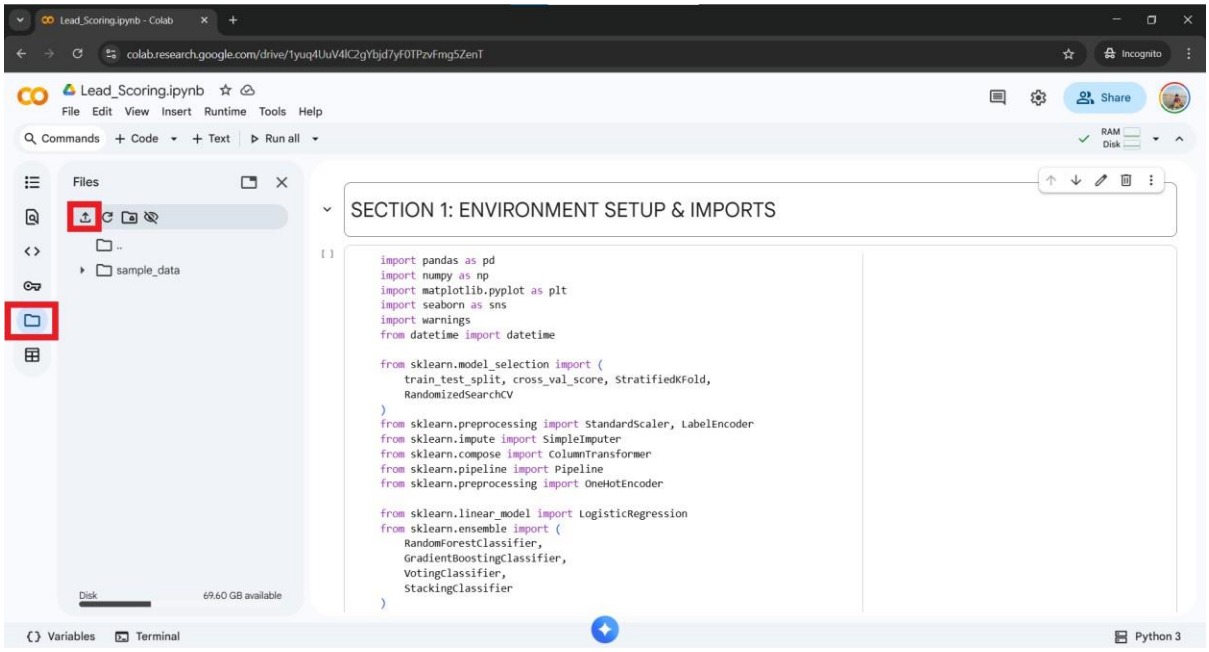
Step 3: Upload the Dataset

1. Important Step: The dataset must be present in the code's runtime environment.

2. Click the Folder icon (Files) on the left sidebar of the Colab interface.

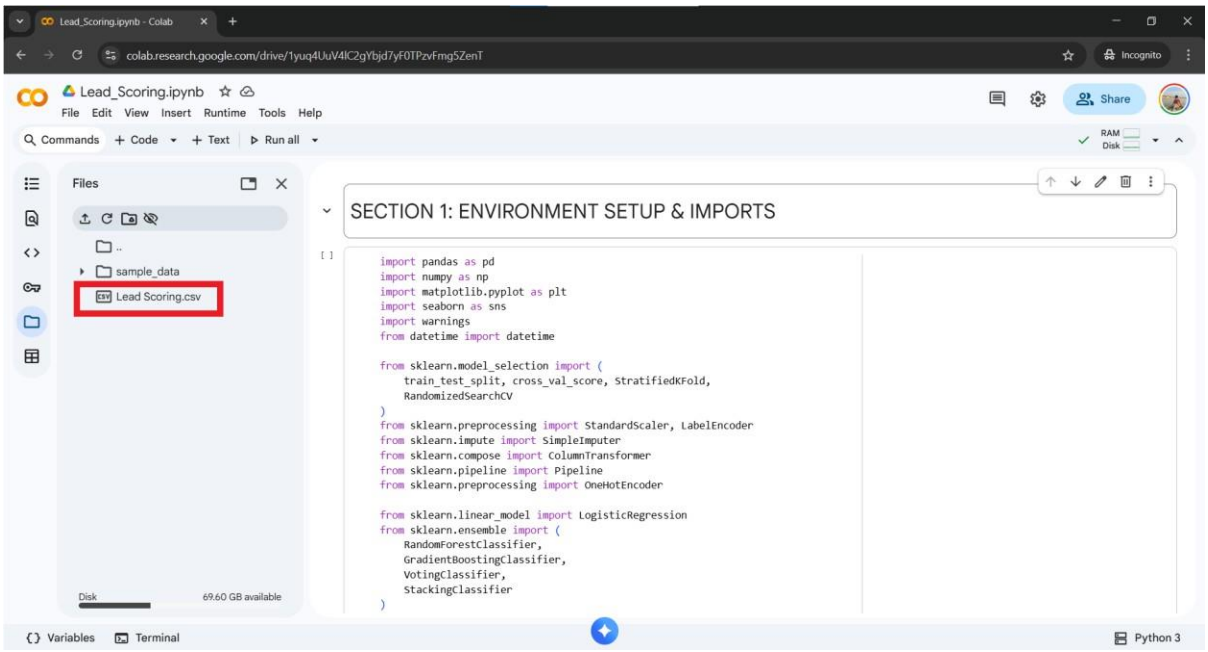


3. Click the Upload button (an arrow pointing up over a piece of paper).

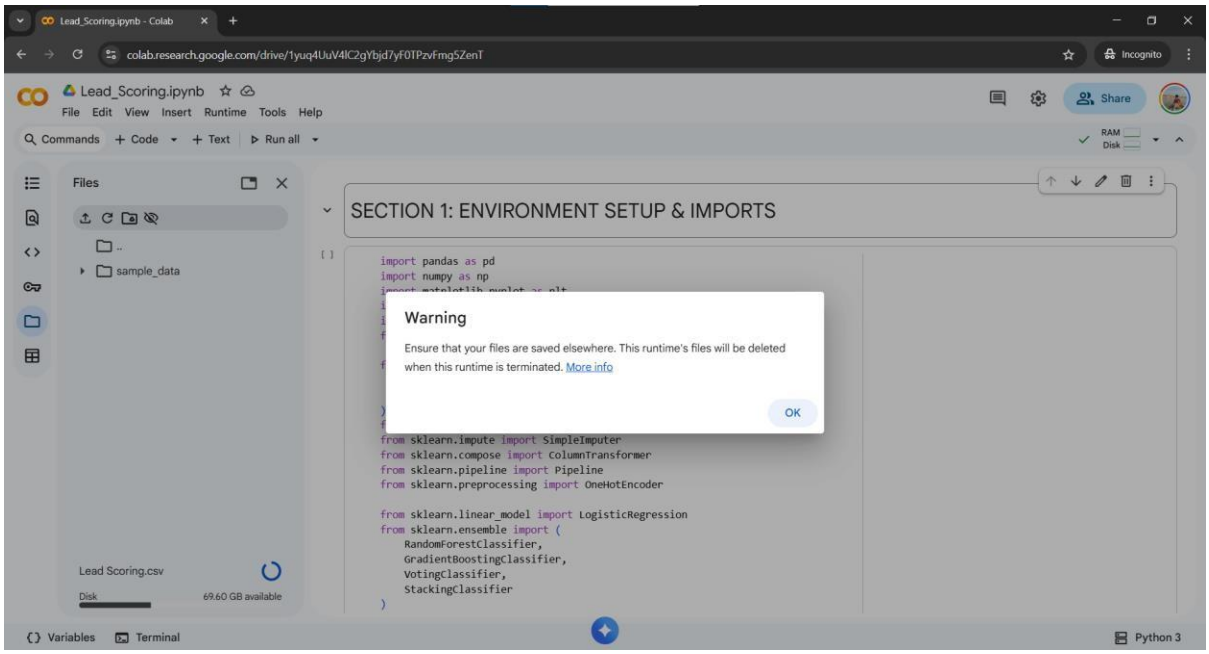


4. Choose the Lead Scoring.csv file from your local machine.

5. Wait until you see the upload circle is finished. The file Lead Scoring.csv should now be listed in the file list.



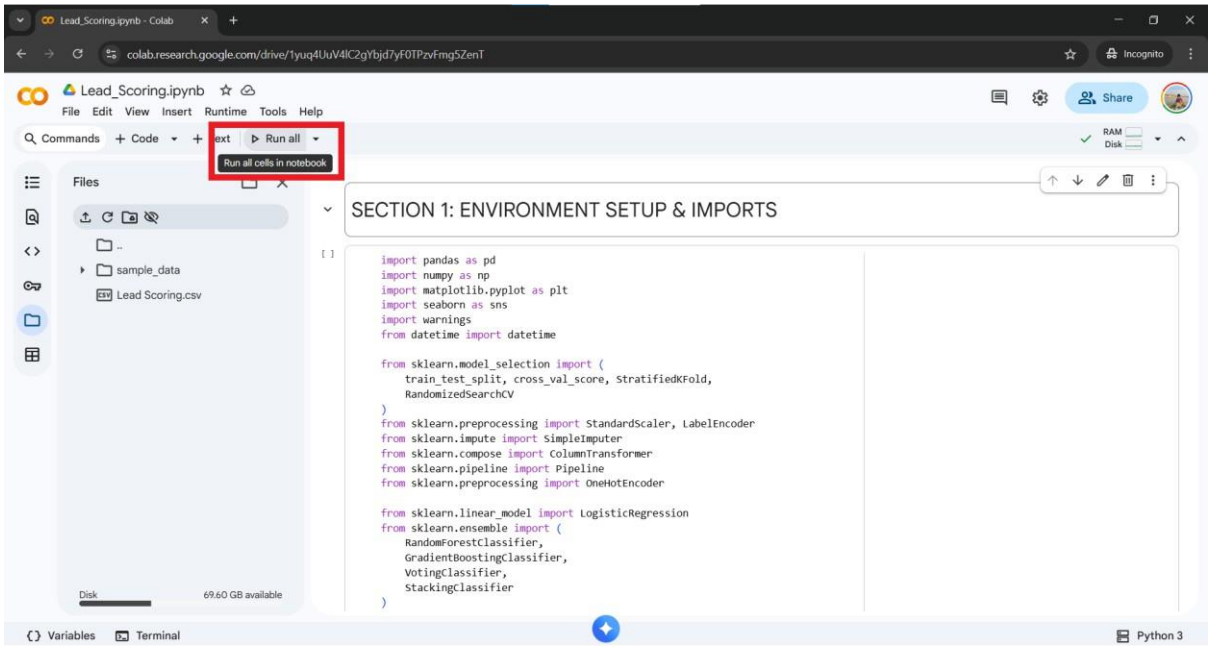
(Note: Files that are uploaded to the Colab session storage are only temporary and will be lost once the runtime recycles.)



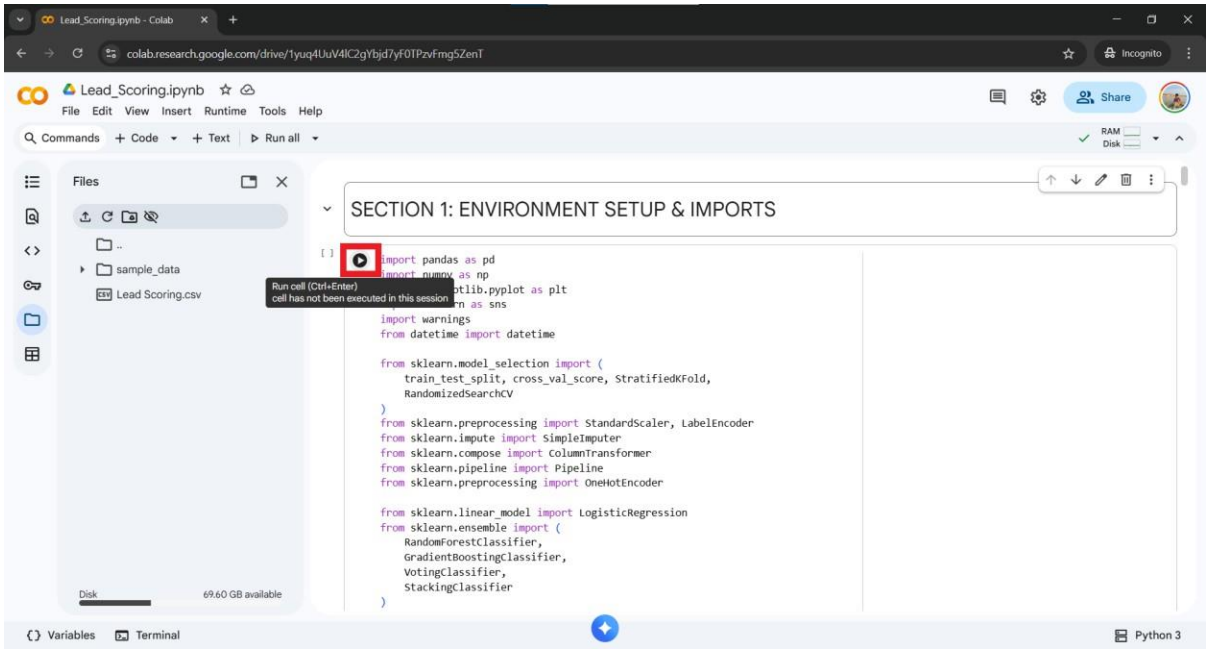
Step 4: Run the Code

1. Open Runtime in the main menu at the top.

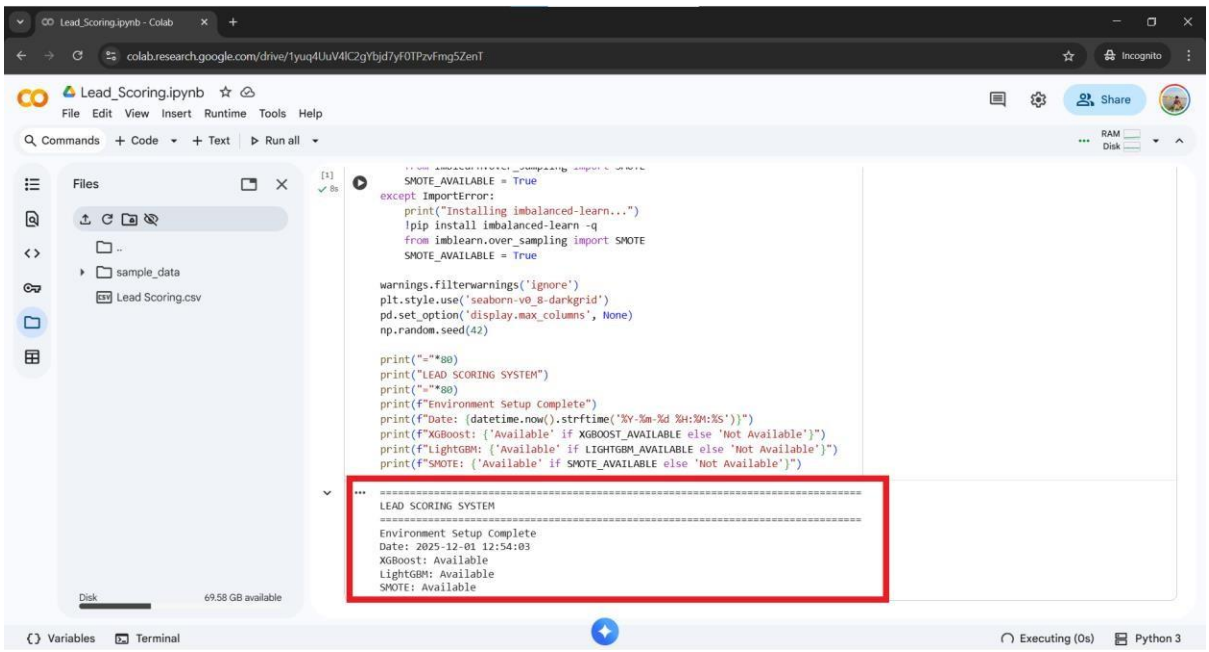
2. Hit Run all (or you can use Ctrl+F9 / Cmd+F9).



3. Another method is by running one code cell at a time by clicking on the Play button, which is on the left side of the respective code cell.



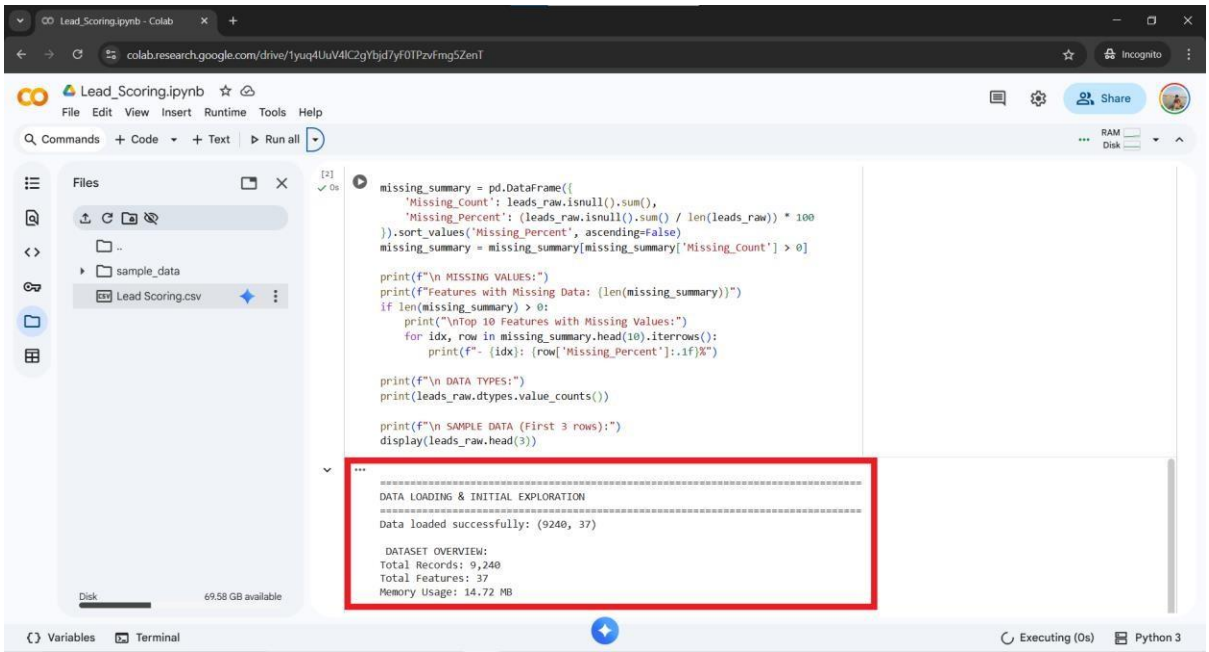
Important Note About Installation: A very first code cell titled "SECTION 1: ENVIRONMENT SETUP" tries to find out if the necessary libraries (xgboost, lightgbm, imblearn) are present or not. In case they are not found, it will resort to a silent method of automatic installation, that is, through !pip install. You might get to witness installation logs in the output.



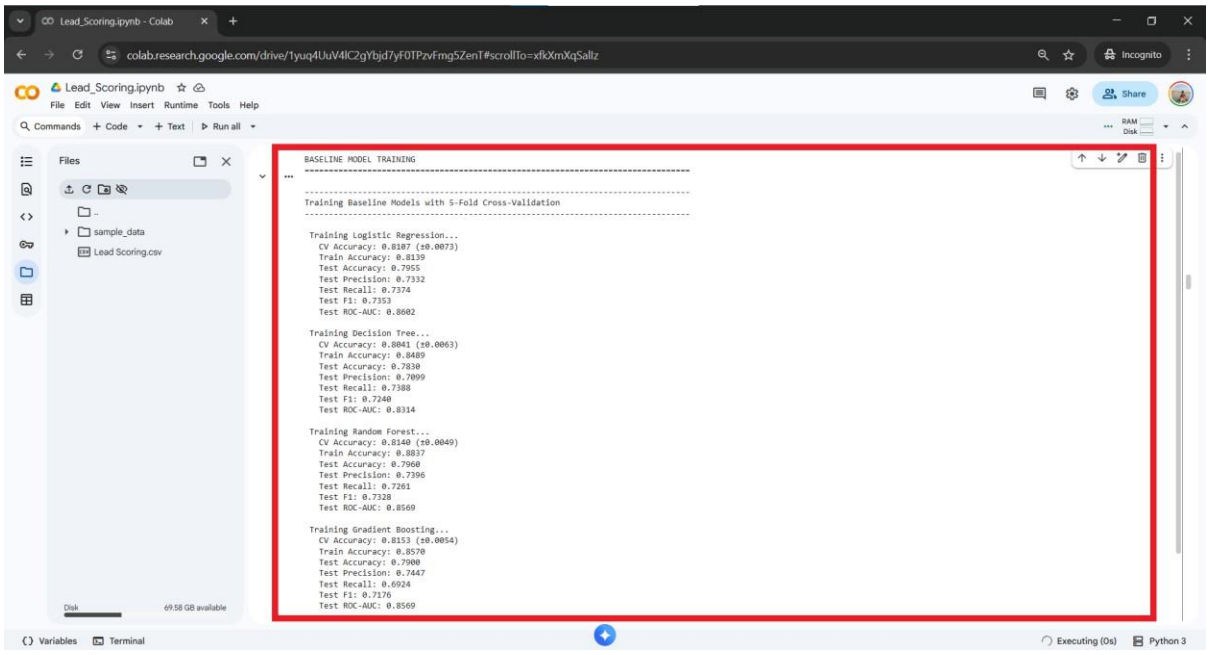
Step 5: Verify Execution

The outputs will appear as you scroll down while the notebook is being executed.

1. Data Loading: The statement "Data loaded successfully: (9240, 37)" should be your confirmation.



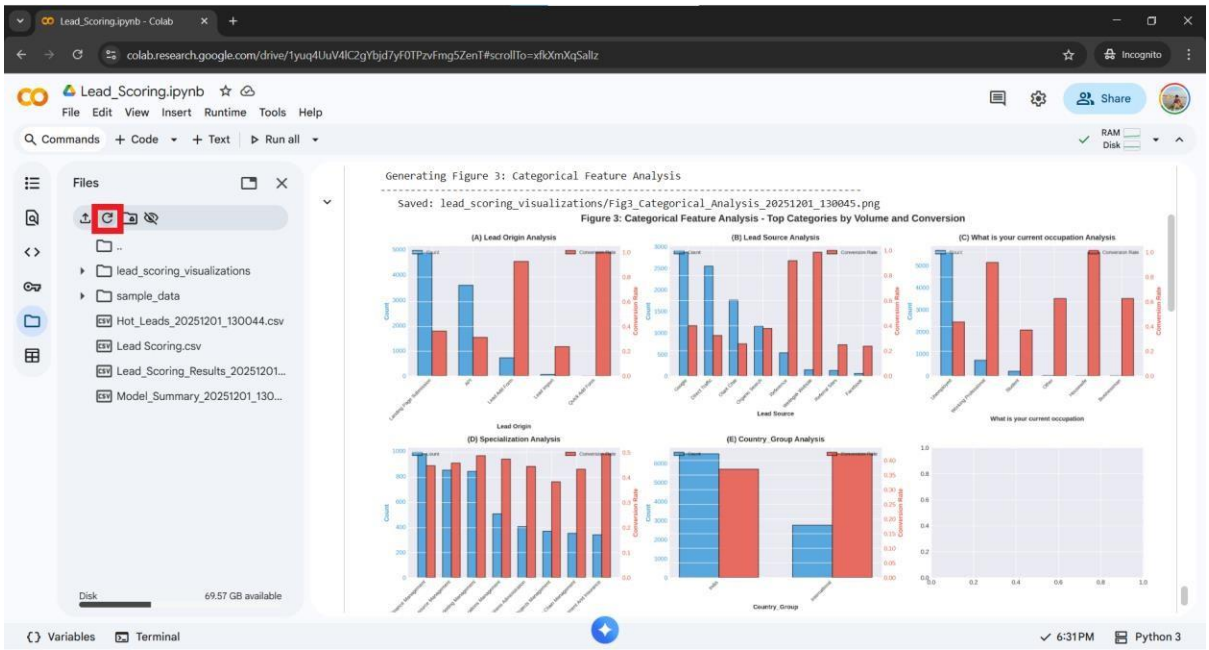
2. Model Training: Watch the progress logs for model training (Logistic Regression, Random Forest, etc.).



3. Visualizations: Plots (histograms, ROC curves, confusion matrices) should appear, and that would be the indication of them being rendered right in the notebook.

5 Output Generation

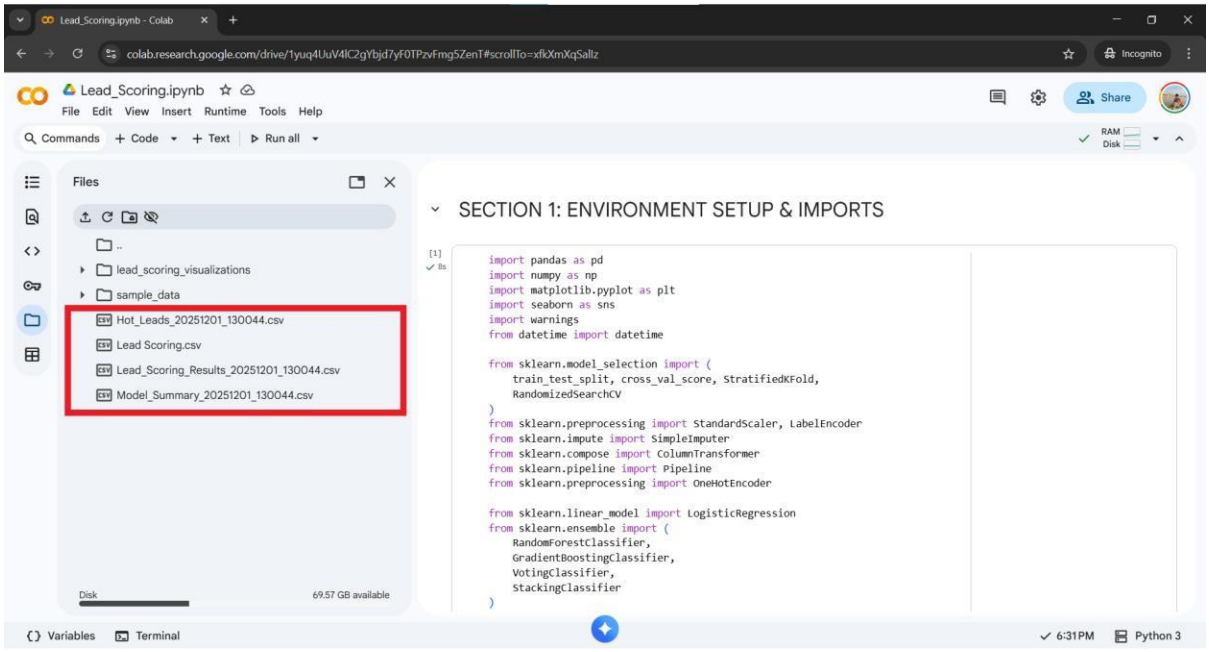
Upon successful execution, the system would produce a bunch of files in the Colab Files pane (which is the left sidebar). You may need to click on the Refresh folder icon to make them appear.



5.1 Generated Files

You have the option of right-clicking any file and then choosing the Download option to get it onto your local machine:

1. Lead_Scoring_Results_[Timestamp].csv: The entire dataset containing predicted probabilities and lead scores (0-100) will be given.
2. Hot_Leads_[Timestamp].csv: A list that is filtered and contains only the high-priority "Hot" leads.
3. Model_Summary_[Timestamp].csv: A document that provides information about the accuracy and metrics of the final model through a report.
4. lead_scoring_visualizations/: A directory that always has the high-resolution PNG images of all the charts that are generated (ROC Curves, Feature Importance, etc.) as part of the visualizations.



References

Google Colab: Google. (n.d.). Google Colab Frequently Asked Questions. Available at: https://research.google.com/colaboratory/faq.html

Python: Python Software Foundation. (n.d.). Python 3.10.12 Documentation. Available at: https://docs.python.org/3.10/

Scikit-learn: Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python. Available at: https://scikit-learn.org/stable/user_guide.html

Pandas: The Pandas Development Team. (2024). pandas documentation. Available at: https://pandas.pydata.org/docs/

NumPy: Harris, C.R. et al. (2020). NumPy Documentation. Available at: https://numpy.org/doc/stable/

XGBoost: Chen, T. and Guestrin, C. (2016). XGBoost Documentation. Available at: https://xgboost.readthedocs.io/en/stable/

Imbalanced-learn: Lemaitre, G., Nogueira, F. and Aridas, C.K. (2017). imbalanced-learn documentation. Available at: https://imbalanced-learn.org/stable/

Matplotlib: Hunter, J.D. (2007). Matplotlib: Visualization with Python. Available at: https://matplotlib.org/stable/index.html