DIGITAL HUB INSIGHTS

Getting Started Guide

GPT-Augmented Momentum Strategy

How to set up your environment, install the notebooks,
and run the course on Google Colab

Version 1.0 | February 2026

Mehrzad Mahdavi, PhD

digitalhub-labs.com

Before You Begin

This course runs entirely in the browser using Google Colab. You do not need to install Python, Jupyter, or any packages on your computer. All you need is:

- A Google account (free) for Google Drive and Google Colab
- A modern web browser (Chrome, Edge, Firefox, or Safari)
- The course files downloaded from your Podia dashboard

> **No Installation Required**
> Google Colab provides a pre-configured Python environment with all major libraries (pandas, scikit-learn, XGBoost, matplotlib) already installed. You can start coding immediately.

Course Overview

The course consists of 8 episodes, each delivered as a Jupyter notebook (.ipynb file). The episodes build on each other sequentially, with data flowing from one episode to the next via Google Drive.

Ep.	Title	Description
00	**Quick Start — Preview of Results**	Preview of final strategy performance and course roadmap
01	**Introduction — The GPT-Augmented Thesis**	Course philosophy, the governance-first approach, and what makes this different
02	**Data Preparation**	Download S&P 500 data, clean and structure for ML pipeline
03	**Feature Engineering**	Create momentum, volatility, and technical features from price data
04	**Model Training**	Train and compare ML models (XGBoost, Ridge, Random Forest, etc.)
05	**Backtesting & Leakage Detection**	Run initial backtest, detect and fix data leakage — the critical episode
06	**Walk-Forward Architecture**	Production-ready validation with rolling retrain windows
07	**Results & Performance Analysis**	Complete metrics: Sharpe, drawdown, risk-adjusted returns

Episode 05 is the critical turning point — where we discover data leakage in the initial backtest and use GPT to debug and fix it. This is the core lesson of the course: governance and honest assessment matter more than impressive numbers.

Step 1: Set Up Google Drive

The notebooks save and load data through Google Drive. Setting up your folder structure takes about two minutes and only needs to be done once.

Create the Course Folder

1. Go to **drive.google.com** and sign in with your Google account.
2. Click New → New Folder and name it exactly: XGBoost_Course
3. Open the XGBoost_Course folder.
4. Inside it, create a subfolder named: Data
5. Create another subfolder named: Notebooks
6. Create a third subfolder named: Results

Your folder structure should look like this:

Folder	Contents	Notes
XGBoost_Course/	Root folder	Top-level course folder
Data/	Market data	CSV file lives here
Notebooks/	All .ipynb files	Episodes 00–07
Results/	Auto-generated	Models, outputs, logs

> **Important: Use Exact Names**
> The notebooks expect these exact folder names. Using different names (e.g., "xgboost" instead of "XGBoost_Course") will cause file-not-found errors.

Step 2: Upload Course Files

From your Podia dashboard, download the course files. Then upload them to your Google Drive folders.

Upload the Data File

1. Locate yf_market_data_2019_2024.csv in your downloaded course files.
2. Upload it to: MyDrive / XGBoost_Course / Data /
3. Verify the file appears in your Data folder.

> **One Upload, All Episodes**

This single CSV file is used by all episodes. Once it is in your Drive, you never need to upload it again — every notebook will find it automatically.

Upload the Notebooks

1. Locate the .ipynb notebook files (Episodes 00 through 07) in your downloaded course files.
2. Upload all notebooks to: MyDrive / XGBoost_Course / Notebooks /
3. Verify all 8 notebooks appear in the Notebooks folder.

After uploading, your Drive should contain:

```
XGBoost_Course/
  Data/
    yf_market_data_2019_2024.csv
  Notebooks/
    00_Quick_Start.ipynb
    01_Introduction.ipynb
    02_Data_Preparation.ipynb
    03_Feature_Engineering.ipynb
    04_Model_Training.ipynb
    05_Backtesting_Framework.ipynb
    06_WalkForward_Architecture.ipynb
    07_Results_Performance_Analysis.ipynb
  Results/
    (empty — populated automatically as you run episodes)
```

Step 3: Open a Notebook in Google Colab

Google Colab is a free, browser-based environment for running Jupyter notebooks. Here is how to open any episode.

1. In Google Drive, navigate to XGBoost_Course / Notebooks.
2. Double-click **any .ipynb file**. Google Drive will open it in Colab automatically.
3. If prompted to choose an app, select Google Colaboratory. (If Colab is not listed, click "Connect more apps" and search for Colaboratory.)
4. The notebook will open in a new browser tab, ready to run.

> **Bookmark Colab**
> You can also open Colab directly at colab.research.google.com, then use File → Open Notebook → Google Drive to browse your notebooks.

Step 4: Connect to Google Drive

Each notebook includes a setup cell that connects to your Google Drive so it can read the data file and save results. This step is required once per session.

1. In the notebook, find the cell labeled "Google Drive Setup" or "Data Source Configuration" near the top.
2. Set **USE_DRIVE = True** (it may default to False).
3. Run the cell (click the play button or press Shift + Enter).
4. A popup will ask you to connect to Google Drive. Click "Connect to Google Drive."
5. Select your Google account and grant permission.
6. Once connected, the cell will confirm the data file was found.

> **Re-connect Each Session**
> Google Colab sessions expire after a period of inactivity (typically 90 minutes idle or 12 hours total). When you start a new session, you will need to re-run the Drive setup cell. This takes about two seconds.

Step 5: Run the Notebook

With Drive connected, you are ready to work through the episode.

Running Cells

- Click on a cell and press Shift + Enter to run it and move to the next cell.
- Alternatively, use Runtime → Run All to execute the entire notebook.
- Cells must be run in order from top to bottom. Skipping cells will cause errors.

Saving Results

Each episode automatically saves its outputs (processed data, trained models, performance logs) to the Results folder in your Drive. The next episode loads from this same folder, creating a seamless pipeline.

Recommended Workflow

- Start with Episode 00 (Quick Start) for a preview of the final results.
- Read Episode 01 (Introduction) for the course thesis and methodology.
- Work through Episodes 02–07 in sequence. Each one builds on the previous.
- Take your time. Read the markdown cells — they explain the reasoning behind every decision.

Troubleshooting

"File not found" Error

- Verify the CSV file is in: MyDrive / XGBoost_Course / Data /
- Check the filename matches exactly: yf_market_data_2019_2024.csv
- Re-run the Drive setup cell and confirm permissions were granted.

"Drive not mounted" Error

- Make sure USE_DRIVE = True in the setup cell.
- Re-run the cell and click "Connect to Google Drive" when prompted.
- If it still fails: Runtime → Disconnect and Delete Runtime, then reconnect.

Session Disconnected or Timed Out

- Colab sessions expire after inactivity. Simply reconnect and re-run the cells from the top.
- Your data and results are safe in Google Drive — nothing is lost.

Notebook Runs Slowly

- Go to Runtime → Change Runtime Type and ensure the hardware accelerator is set to "None" or "CPU." This course does not require GPU.
- If Colab is throttling your usage, wait a few minutes and try again.

"Permission denied" Error

- Go to Runtime → Disconnect and Delete Runtime.
- Reconnect and re-run all cells from the top, granting Drive permissions when prompted.

Quick Reference

Action	How
Open Colab	`colab.research.google.com`
Open notebook from Drive	Double-click .ipynb file in Google Drive
Run a cell	Shift + Enter
Run all cells	Runtime → Run All

Restart runtime	Runtime → Restart Runtime
Reset everything	Runtime → Disconnect and Delete Runtime
Check GPU setting	Runtime → Change Runtime Type
Mount Drive manually	`from google.colab import drive;` `drive.mount('/content/drive')`

Questions? Reach out at digitalhub-labs.com
Happy learning!