

## MSDS 451 - Coding Assignment 3

Ethan Vanlerberghe

School of Professional Studies, Northwestern University

MSDS 451: Financial Engineering

Dr. Miller

31 August 2025



NORTHWESTERN  
UNIVERSITY

**Introduction (10 points). Why are you conducting this research? Identify potential users of the knowledge base and application(s) that you intend to develop.**

The purpose of conducting this research is to create an algorithmic trading model to buy and hold various assets based on a momentum trading strategy. More specifically, my project evaluates the length of review that the model used to predict whether or not to purchase the asset. To do so, thorough research must be performed to ensure proper decision-making can be made every step of the way, leading to the best possible outcome for any future use. This research, however, can extend beyond just the creation of this ETF, and instead help others who are looking to set up their "Own ETF's", as it provides a strong basis of understanding for those beginning to understand the purpose of ETFs and how to create one successfully. This research should be used by individuals looking to create their "Own ETF's", to establish and understand how to select items to build their ETF successfully. Specifically, it should be used to start learning how to build a basic model on 25 years of large-cap fund data. The end application is designed to be used by others as a teaching tool to better see the possibilities of creating their "Own ETFs" and allow users to track the performance and decision-making of the model. Specifically, this should be used by any individuals who are learning to create algorithmic trading platforms for the first time.

**Literature review (10 points).. Who else has conducted research like this?**

A great deal of research has already been done in this area, and serves as a massive knowledge base on how to begin this process. Key examples lie in GitHub projects (AlpacaHQ, 2019), or Youtube videos (Algovibes, 2025). In these examples, assets are pulled from Yahoo Finance using YFinance. Then, they make trades based on past performance of the asset. In the AlpacaHQ example, the model tracks an asset's performance over the previous day was  $>4\%$ , while also checking asset volume to ensure that it is true performance. This type of example follows the same trend that the model this paper discusses follows.

**Methods (10 points). How are you conducting the research? Make sure you address the issues that are the focus of this checkpoint assignment.**

The project began by running the example code directly to see an AUC of 0.54 and a confusion matrix showing a 46% true positive and 31% true negative rate (See Fig. 1). Reading into how the model was developed, it was identified that the prediction of the model was based on a very short (1-day) momentum strategy. Here, if the current day closed higher than the previous day, it would flag a 1, and if not, flag a 0.

In order to continue with the momentum-based strategy, an initial extension was created, allowing the momentum model to look over the course of 2 days, and predict based on that. If the asset trended positively all three days, it would flag a 1, if not, it would flag a 0. After hyperparameter tuning, this model turned to perform slightly better than the first. This resulted in an AUC of 0.50, a true positive rate of 45%, and a true negative rate of 40% (See Fig. 2). Looking at the chart of feature importance, all of the “Up” indicators have jumped up multiple times, with vb, gld, and spy as the top 3. Due to this success, one more attempt was made with a third lagging day.

A third model was then created, following this same trend, however utilizing a momentum strategy over 3 days. This, meaning that the model would observe an asset’s value over 3 days. If it followed the same progression, it would yield a 1, if not, a 0. This yielded an AUC of 0.55, a true positive of 29%, but a true negative of 60% (See Fig. 3). This means that the model was much more successful at finding downward trajectories, and poor at tracking upward ones.

**Conclusions (10 points).. So, what does it all mean? Do you have any concerns about the term project at this point?**

Overall, I learned quite a bit about how algorithmic trading operates. While the main thing that I altered was the momentum strategy itself, studying how the data was brought in, manipulated, normalized and then the model built was where I learned the most. I was surprised by how the accuracy scoring

operates against the estimator to project the test accuracy scores. Visualizing this through the AUC curve and Confusion matrix really put it into context.

Figure 1: Portfolio Performance (Jumpstart)

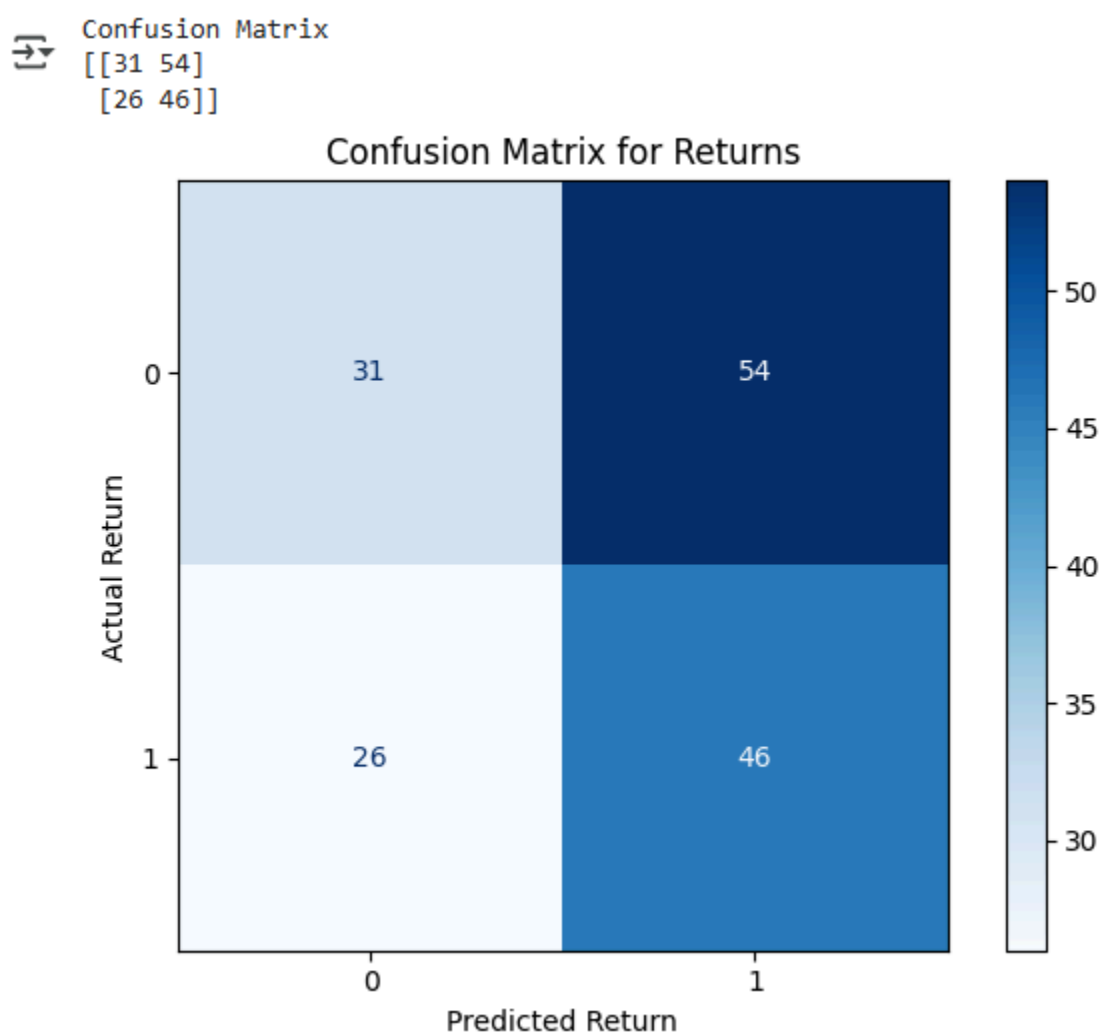


Figure 2: Portfolio Performance (Lagging 2 day)

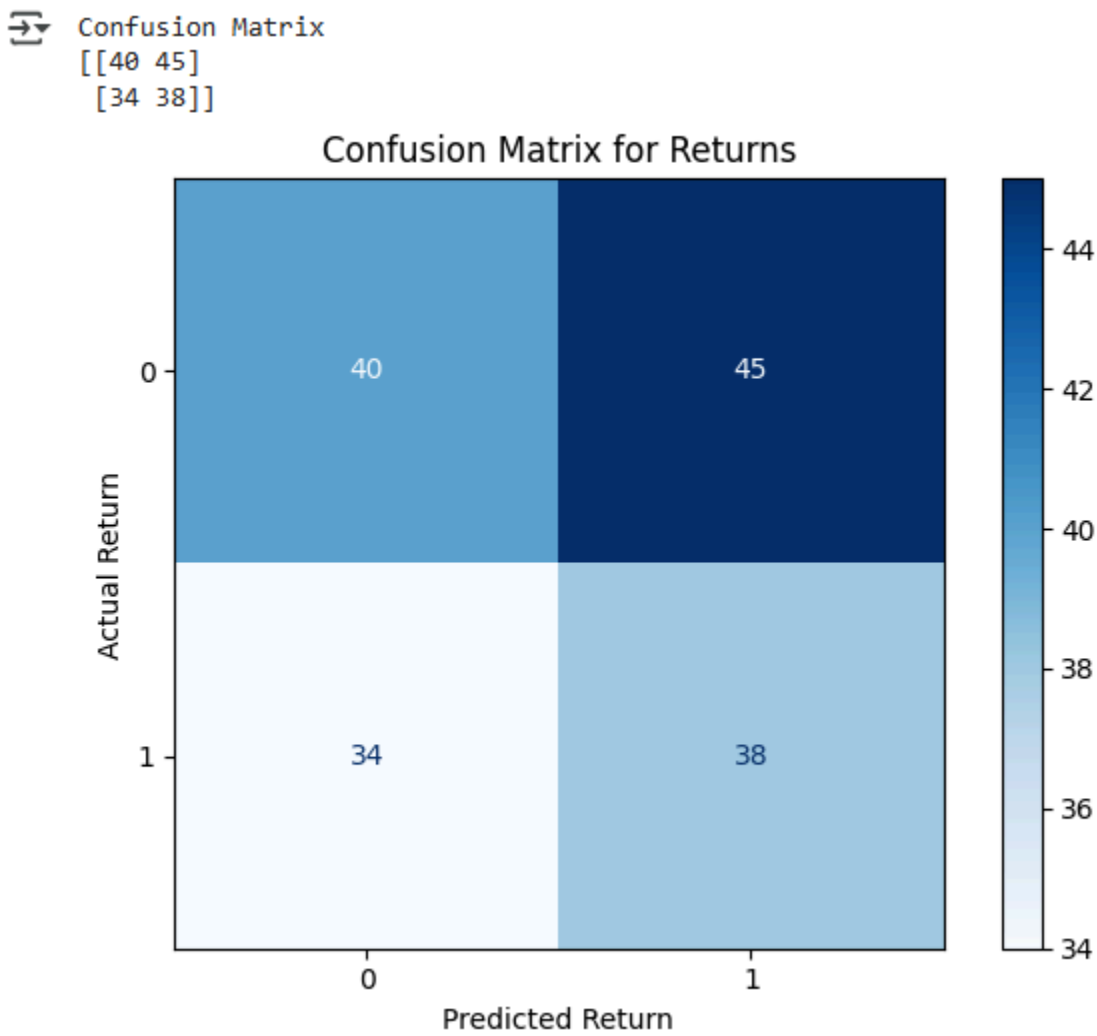
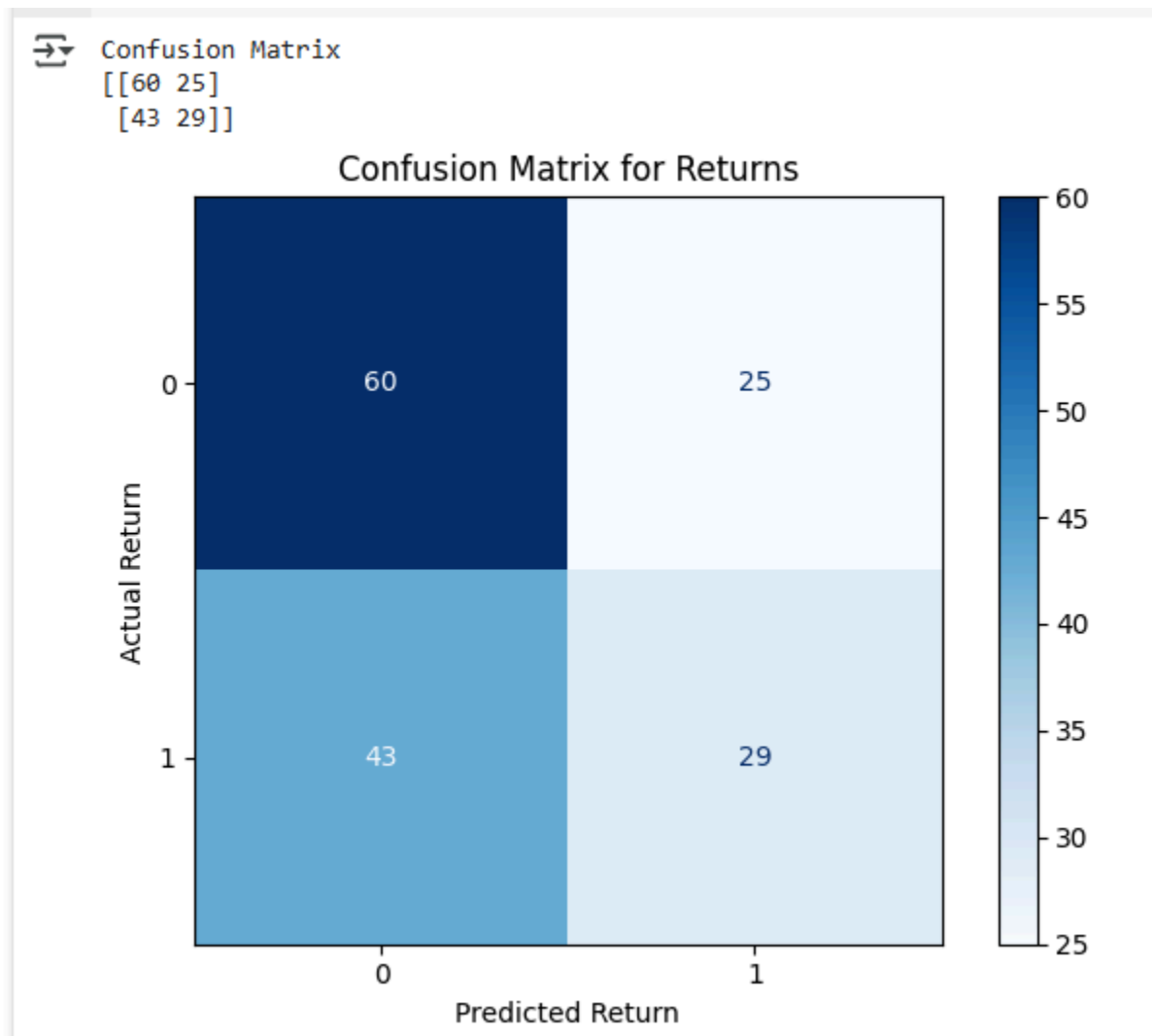


Figure 3: Portfolio Performance (Lagging 3 day)



## Works Cited

AlpacaHQ. *Momentum-Trading-Example*. GitHub, 13 Aug. 2025,  
github.com/alpacaHQ/Momentum-Trading-Example. Accessed 25 Aug. 2025.

*Master Momentum Trading with Python: Full Time-Series Strategy ...* YouTube, uploaded by [Uploader's  
Name], Jan. 2025, [www.youtube.com/watch?v=dC1Cl2E-lxg](https://www.youtube.com/watch?v=dC1Cl2E-lxg). Accessed 25 Aug. 2025.

Miller, Tom. "451 Feature Engineering: Programming Assignment 3." 2025.

—. "Week 7 Programming Assignment 3." Lecture, 2025.