10 Academy: Artificial Intelligence Mastery

Week 11 Challenge Document

Date: 06 Aug - 12 Aug 2025

Time Series Forecasting for Portfolio Management Optimization

Business objective

Guide Me in Finance (GMF) Investments is a forward-thinking financial advisory firm that specializes in personalized portfolio management. GMF leverages cutting-edge technology and data-driven insights to provide clients with tailored investment strategies. By integrating advanced time series forecasting models, GMF aims to predict market trends, optimize asset allocation, and enhance portfolio performance. The company's goal is to help clients achieve their financial objectives by minimizing risks and capitalizing on market opportunities.

The Efficient Market Hypothesis suggests that predicting exact stock prices using only historical price data is exceptionally difficult. Therefore, in an industry setting, these models are more often used to forecast volatility, identify momentum factors, or serve as one of many inputs into a larger decision-making framework, rather than for direct, standalone price prediction.

At GMF Investments, financial analysts play a crucial role in interpreting complex financial data and providing actionable insights. By utilizing real-time financial data from sources like YFinance, GMF ensures its strategies are based on the latest market conditions, thereby maintaining a competitive edge.

Situational Overview (Business Need)

As a **Financial Analyst** at GMF Investments, your objective is to apply time series forecasting to historical financial data to enhance portfolio management strategies. Your role involves analyzing data, building predictive models, and recommending portfolio adjustments based on forecasted trends.

You will:

- Utilize **YFinance** data to extract historical financial information such as stock prices, market indices, and other relevant financial metrics.
- Preprocess and analyze this data to identify trends and patterns.
- Develop and evaluate forecasting models to predict future market movements.
- Use the insights gained to recommend changes to client portfolios that aim to optimize returns while managing risks.

Data

Use historical financial data for three key assets: Tesla (TSLA) Historical stock prices (Open, High, Low, Close), volume, and volatility., Vanguard Total Bond Market ETF (BND), and S&P 500 ETF (SPY). The data will be sourced from **YFinance** and cover the period from **July 1, 2015, to July 31, 2025**.

Fetch the data using the **YFinance** Python library **July 1, 2015, to July 31, 2025.**

Each dataset includes:

- Date: Trading day timestamp.
- Open, High, Low, Close: Daily price metrics, with Adj Close representing the adjusted close price to account for dividends and splits.
- Volume: The total number of shares/units traded each day.

Asset-Specific Descriptions

- TSLA: High-growth, high-risk stock in the consumer discretionary sector (Automobile Manufacturing).
- BND: A bond ETF tracking U.S. investment-grade bonds, providing stability and income.
- SPY: An ETF tracking the S&P 500 Index, offering broad U.S. market exposure.

Usage in Portfolio Analysis

- TSLA provides potential high returns with high volatility.
- BND contributes stability and low risk.
- SPY offers diversified, moderate-risk market exposure.

Expected Outcomes:

Skills:

- API Usage: Skillfully fetching financial data from an API (yfinance).
- Data Wrangling: Using pandas for cleaning, handling missing dates/values, time-based indexing, and merging datasets.
- Feature Engineering: Calculating daily returns, rolling volatility, and other relevant metrics.
- Data Scaling: Applying normalization or standardization as a preprocessing
- Statistical Modeling: Building, training, and optimizing an ARIMA/SARIMA model using libraries like statsmodels and pmdarima
- Deep Learning Modeling: Constructing, training, and evaluating an LSTM model for time series forecasting
- Model Evaluation: Calculating and comparing performance metrics
- Optimization & Visualization: Skillfully running simulations to generate and plot the Efficient Frontier and identifying key portfolios
- MPT Implementation: Using libraries like PyPortfolioOpt
- Simulation: Implementing a simple backtesting loop to simulate portfolio performance over a historical period.
- Professional Communication

Knowledge:

- Understanding the characteristics of different asset classes: high-growth stocks (TSLA), bonds for stability (BND), and market indices for diversification (SPY).
- Familiarity with the Efficient Market Hypothesis (EMH) and its practical implication that pure price prediction is difficult.
- Deeply understanding what stationarity is, why it's crucial for models like ARIMA, and how to test for it
- Knowing what the frontier represents and the significance of portfolios that lie on it.
- Understanding the purpose and methodology of backtesting a financial strategy.
- Knowing the importance of using a benchmark for objective performance evaluation.

Abilities

- Critical Evaluation: The ability to compare and contrast different modeling approaches
- Problem Framing & Synthesis: The ability to translate a high-level business objective
- Data-Driven Decision Making

Team

Tutors:

- Mahlet
- Rediet
- Kerod
- Rehmet

Key Dates

- Discussion on the case - Wednesday 06 Aug 2025. Use #all-week11 to pre-ask questions.
- Interim Solution - 20:00 UTC on Sunday 10 Aug 2025.
- Final Submission - 20:00 UTC on Tuesday 12 Aug 2025

Instructions

Objectives:

The objective of this challenge is to equip trainees with the skills to preprocess financial data, develop time series forecasting models, analyze market trends, and optimize investment portfolios. Participants will gain hands-on experience in leveraging data-driven insights to enhance portfolio performance, minimize risks, and capitalize on market opportunities.

Task 1: Preprocess and Explore the Data

Load, clean, and understand the data to prepare it for modeling.

- Extract historical financial data using YFinance for :
    a. TSLA provides potential high returns with high volatility.
    b. BND contributes stability and low risk.
    c. SPY offers diversified, moderate-risk market exposure.
- Data cleaning and Understanding.
    a. Check basic statistics to understand the distribution of the data.
    b. Ensure all columns have appropriate data types and check for missing values.
    c. Handle missing values by either filling, interpolating, or removing them.
    d. Normalize or scale the data if required, especially for machine learning models.
- Conduct Exploratory Data Analysis (EDA):
    a. Visualize the closing price over time to identify trends and patterns.
    b. Calculate and plot the daily percentage change to observe volatility.
    c. Analyze volatility by calculating rolling means and standard deviations to understand short-term trends and fluctuations.
    d. Perform outlier detection to identify significant anomalies.
        i. Analyze days with unusually high or low returns.
- Seasonality and Trends:
    a. Perform a statistical test (e.g., Augmented Dickey-Fuller test) on the closing prices and daily returns. Discuss the results and their implications. A non-stationary series requires differencing (the 'I' in ARIMA) to become stationary, which is a prerequisite for the model.
- Analyze Volatility
    a. Calculate rolling means and standard deviations to understand short-term trends and volatility.
- Document key insights like overall direction of Tesla's stock price, Fluctuations in daily returns and their impact, and calculate foundational risk metrics like Value at Risk (VaR) and the Sharpe Ratio to assess potential losses and historical risk-adjusted returns.

Task 2: Develop Time Series Forecasting Models

This task involves building a time series forecasting model to predict Tesla's future stock prices. Below are the step-by-step instructions to develop, evaluate, and refine a forecasting model using common techniques such as ARIMA, SARIMA, or LSTM.

- Implement and compare at least two different types of models
  - **A classical statistical model:** ARIMA (AutoRegressive Integrated Moving Average) or SARIMA (Seasonal ARIMA).
  - **A deep learning model:** LSTM (Long Short-Term Memory).
  - This comparison will allow you to analyze the trade-offs between model complexity, performance, and interpretability.
- Divide the dataset into training and testing sets. Crucially, the data must be split chronologically to preserve the temporal order (e.g., train on 2015-2023, test on 2024-2025). Random shuffling is inappropriate for time series data.
- Use the models to forecast future stock prices and compare the predictions with the test set.
- **Optimize Model Parameters**:
  - Use techniques like grid search or auto_arima to find the best (p, d, q) parameters for ARIMA. For LSTM, experiment with architecture (layers, neurons) and hyperparameters (epochs, batch size).
- Use the models to forecast over the test set period. Compare the performance of all implemented models using metrics like MAE, RMSE, and MAPE. Provide a brief discussion on which model performed better and why that might be the case.

Task 3: Forecast Future Market Trends

In this task, you'll use the model developed in Task 2 to forecast Tesla's future stock prices. The goal is to generate future price predictions, analyze the results, and provide insights on potential trends and risks.

- Use the Trained Model for Forecasting
  - Depending on the model you chose (ARIMA, SARIMA, or LSTM), you'll generate forecasts for 6-12 months.
- Forecast Analysis
  - Visualize the forecast alongside historical data.
  - The forecast should include confidence intervals to show the range within which the future prices are expected to lie.
- **Interpret the Results**
  1. **Trend Analysis**:
     - Look for long-term trends (upward, downward, or stable).

    - Identify any patterns or anomalies in the forecast.
2. **Volatility and Risk**:
    - Discuss the level of uncertainty captured by the confidence intervals.
    - Critically analyze the confidence intervals. How does their width change over the 6-12 month forecast horizon? What does this imply about the reliability and certainty of long-term forecasts?
3. **Market Opportunities and Risks**:
    - Based on the forecast, outline potential market opportunities (e.g., expected price increases) and risks (e.g., high volatility or expected declines).

Task 4: Optimize Portfolio Based on Forecast

In this task, you'll use insights from your forecast to make informed decisions about portfolio construction, grounded in the principles of Modern Portfolio Theory (MPT).

- **Forecasted Asset (TSLA**): Use the return forecast generated by your best-performing model from Task 2 as the expected return for Tesla.
- **Historical Assets (BND, SPY)**: For the more stable assets, BND and SPY, we will use their historical average daily returns (annualized) as the proxy for their expected returns. This simulates a common approach where an analyst has a specific "view" on one asset while relying on historical data for others.
- **Covariance Matrix**: Compute the covariance matrix based on the historical daily returns of all three assets (TSLA, BND, SPY). This matrix is crucial for understanding how the assets move together and for calculating portfolio risk.
- Using the expected returns vector and the covariance matrix, run an optimization simulation to generate the Efficient Frontier. This frontier represents the set of optimal portfolios that offer the highest expected return for a defined level of risk.
- Plot the Efficient Frontier with portfolio volatility (risk) on the x-axis and portfolio return on the y-axis.
- On your plot, identify and mark two key portfolios:
    - The Maximum Sharpe Ratio Portfolio (aka The Tangency Portfolio).
    - The Minimum Volatility Portfolio.
- Based on your analysis of the Efficient Frontier, select and recommend an optimal portfolio. Justify your choice (e.g., are you prioritizing maximum risk-adjusted return, or are you aiming for lower risk?).
- Summarize your final recommended portfolio, including:
    - The optimal weights for TSLA, BND, and SPY.
    - The portfolio's expected annual return, volatility, and Sharpe Ratio.

Task 5: Strategy Backtesting

A forecast and an optimized portfolio are hypotheses. A backtest is the experiment that validates a strategy by simulating its performance on historical data. In this task, you will simulate the performance of your proposed strategy and compare it against a simple benchmark.

Instructions:

1. **Define a Backtesting Period:** Use the last year of your dataset (e.g., August 1, 2024 - July 31, 2025) as your backtesting window.
2. **Define a Benchmark:** Create a simple benchmark portfolio to compare against, such as a static 60% SPY / 40% BND portfolio.
3. **Simulate Your Strategy:**
    - Start with the initial optimal weights you found in Task 4.
    - "Hold" this portfolio for a set period (e.g., one month).
    - At the end of the month, you would ideally re-run your forecast and re-optimize. For simplicity, you can either hold the initial weights for the full year or perform a simplified rebalancing.
4. **Analyze Performance:**
    - Plot the cumulative returns of your strategy portfolio against the benchmark portfolio over the backtesting period.
    - Calculate the final Sharpe Ratio and total return for both your strategy and the benchmark.
    - Conclude with a brief summary: Did your strategy outperform the benchmark? What does this initial backtest suggest about the viability of your model-driven approach?

Tutorials Schedule

In the following, the color **purple** indicates morning sessions, and **blue** indicates afternoon sessions.

Wednesday

- Introduction to the challenge (Mahlet)
- Comparing time series modeling (Rediet)

Thursday

- Time Series Forecasting and Portfolio Optimization (Kerod)
- Backtesting and Simulation for Trading Strategies (Rehmet)

Friday

- Integrating robust risk analysis into portfolio management(Rediet)

Interim Submission

- Interim report - Covering task - 1
- Link to your GitHub code.

Feedback

You may not receive detailed comments on your interim submission but will receive a grade.

Final Submission

- A professional PDF report framed as an "Investment Memo" for GMF's investment committee, detailing your methodology, findings, and final recommendation. Alternatively, a detailed technical blog post (e.g., on Medium) is also acceptable.
- Link to your Github code, and make sure to screenshots demonstrating anything else you have done.

Feedback

You will receive comments/feedback in addition to a grade.

References

Data Science Workflow

1. https://www.datacamp.com/tutorial/arima
2. https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/
3. https://www.geeksforgeeks.org/time-series-analysis-and-forecasting/
4. https://www.datasciencewithmarco.com/blog/the-complete-guide-to-time-series-forecasting-using-sklearn-pandas-and-numpy
5. https://www.machinelearningplus.com/time-series/time-series-analysis-python/

Portfolio Optimization

1. https://miltonfmr.com/the-complete-guide-to-portfolio-optimization-in-r-part1/
2. https://www.machinelearningplus.com/machine-learning/portfolio-optimization-python-example/
3. https://builtin.com/data-science/portfolio-optimization-python
4. https://www.analyticsvidhya.com/blog/2021/04/portfolio-optimization-using-mpt-in-python/
5. https://github.com/robertmartin8/PyPortfolioOpt
6. https://blog.jetbrains.com/datalore/2024/01/26/portfolio-optimization-in-python-with-datalore-and-ai-assistant/
7. https://tradewithpython.com/building-an-optimal-portfolio-with-python

Efficient Market Hypothesis

1. https://www.investopedia.com/terms/e/efficientmarkethypothesis.asp
2. https://www.investopedia.com/terms/a/adaptive-market-hypothesis.asp
3. https://www.investopedia.com/terms/b/behavioralfinance.asp
4. https://corporatefinanceinstitute.com/resources/wealth-management/asset-class/