

Report

The Portfolio Optimization project aims to enhance the existing portfolio management technique by incorporating an additional layer into the asset selection data. This new layer involves a **1-year projection** of the assets available for selection that will be appended to a database of 2-year historical data used for asset selection. Through this innovative methodology, we anticipate achieving a more robust portfolio that outperforms current solutions both in **higher return** and **lower volatility** or, in other words, **maximizing the Sharpe Ratio**.

This presents an excellent opportunity to explore novel Portfolio Management methodologies that integrate cutting-edge data science techniques currently not prevalent in the industry. While each management fund may have its unique variations of the process, the industry standard primarily relies on historical data for asset selection. By embracing innovative data science approaches, we can potentially gain a competitive advantage and redefine the landscape of portfolio management practices.

The dataset used for this project combines data from two sources: Yahoo Finance and Federal Reserve Economic Data (FRED). The data was directly collected using Python libraries, eliminating the need to upload additional files. The selected timeframe spans from 08-02-2010 to 07-02-2023 on a daily basis. The rationale behind choosing this starting date is to align with a turning-point at the end of the training data, enabling the evaluation of model performance in predicting changes in market directions.

From Yahoo Finance, the following data was collected:

- I. S&P500 index values
- II. S&P500 values for each of the 500 stocks
- III. Dow Jones Industrial index values
- IV. Volatility Index (VIX) values

From FRED, the following data was collected:

- I. Inflation values
- II. Interest Rates values
- III. Long-term vs Short-Term Spread values

All the data sources share the same timeframe and date alignment as the indices, facilitating concatenation and analysis.

During the data preprocessing phase, the following steps were performed:

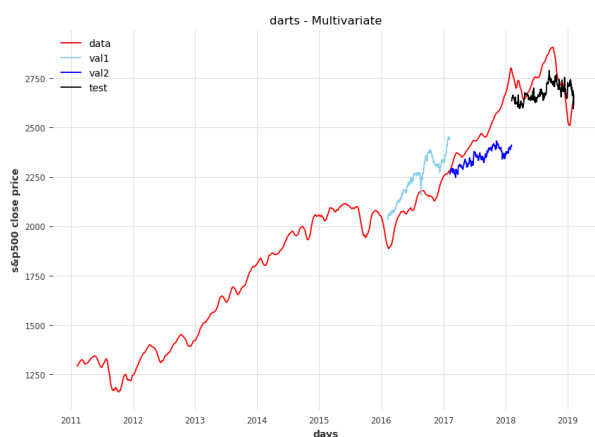
1. Drop Columns: Unnecessary columns like location and open price were removed from all databases.
2. Fill Forward: To account for weekends and holidays when financial information is not available, a fill-forward method was used to propagate the last available value to the missing days.
3. Inflation Monthly to Daily: The inflation database, originally provided on a monthly basis, was transformed to daily frequency, likely by linear interpolation or other suitable methods.
4. Date Time Column: The date format was standardized across all databases to ensure consistency.
5. Feature Engineering: New columns such as "timing" and "30 days moving average for S&P500 index".
6. Removed Leap Years: Some models encountered issues with varying day numbers in leap years, so leap years were excluded from the dataset.
7. Kept Leap Year: Conversely, some models required leap days to be present, so leap years were retained in the dataset.

After the preprocessing steps, the final shape of the dataset used for training, validation 1, validation 2, and test was 4,744 rows by 406 columns. There are 4 more years of data (2019, 2020, 2021, and 2022) that were posteriorly used to analyzed the performance of this methodology.

The entire process involved 5 key steps:

Step 1: Database Consolidation	<ul style="list-style-type: none">• Databases ensemble: 1) S&P500 Index, 2)Dow Jones Ind, 3)Interest Rates, 4)Spread, 5)Inflation, 6) VIX, 7) S&P500 stocks• Additionally, some Feature Engineering and data preparation/cleaning• Result: Database and Database(<u>alldays</u>)
Step 2: Model Selection	<ul style="list-style-type: none">• Models tested: 1)<u>SKLearn</u> Linear Regression, 2)<u>SARIMAX</u>, 3)<u>Prophet</u>, 4) Vector Autoregressive, 5) NN, 6) LSTM, 7) Merlion, 8)Darts• Result: Darts was the model chosen
Step 3: Assets Forecast	<ul style="list-style-type: none">• Forecast of the 500 stocks in the S&P500• Result: <u>Historical.csv</u> (2 years) and <u>Forecast.csv</u> (2 years of historical + 1 year of projections)
Step 4: Assets Selection	<ul style="list-style-type: none">• By using the SLSQP optimizer: Selection of a combination of 30 stocks with that generates the highest Sharpe Ratio• Result: <u>Selected_Stocks(Forecast).csv</u> and <u>Selected_Stock(Historical).csv</u>
Step 5: Portfolio Evaluation	<ul style="list-style-type: none">• Assuming we have invested \$1,000 in the recommended portfolio, analyze the performance for the year 2019, 2020, 2021, and 2022.• Result: Return and Volatility for the Forecast Portfolio, Historical Portfolio, <u>Max_Stock Forecast</u>, <u>Max_Stock Historical</u>, S&P500 Index.

During the Model Selection step, an extensive evaluation of over 10 different models was conducted. However, the focus in the notebooks and the presentation was on the 8 models that exhibited superior performance. Among these 8 models, the Darts Multivariate model stood out as the chosen one due to its remarkable characteristics, such as the smallest accumulated Mean Absolute Error and exceptional performance in predicting turning-points.



Mean Absolute Error (MAE)

Prediction	SKL - LR	SARIMAX	Prophet	VAM	NN	LSTM	Merlion	Darts Mult
1	103	2,079	119	75	97	174	1,105	132
2	262	2,223	171	103	188	314	1,261	133
3	481	24,223	138	178	358	1,178	1,329	93
Total	846	28,525	428	356	643	1,666	3,695	358

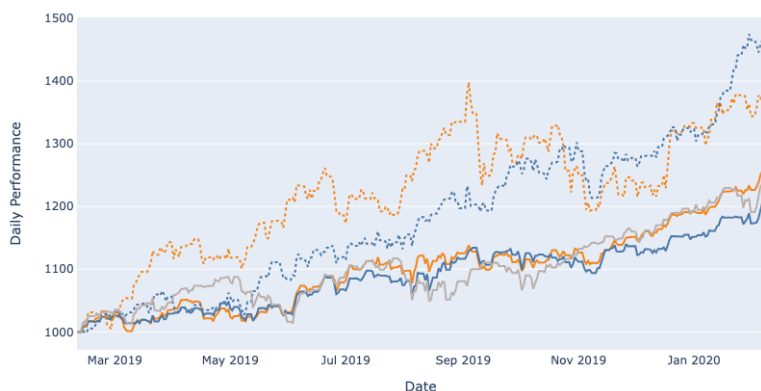
After selecting the Darts Multivariate model, we proceeded with forecasting 1-year stock values for the 500 stocks within the S&P500 index. Subsequently, we sought to identify the optimal 30-stocks portfolio that maximizes the Sharpe Ratio. To evaluate the performance of the **Forecasted Portfolio**, we compared it with the following benchmarks for the years 2019, 2020, 2021, and 2022:

1. **Portfolio historical** (traditional method).
2. **Maximum Stock** from **forecast**.
3. **Maximum Stock** from **historical** data.
4. Performance of the **S&P500 index**.

This comprehensive analysis allowed us to assess the effectiveness and superiority of the **Forecasted Portfolio** against various benchmarks and traditional methods over the specified years.

Results:

Portfolio Daily Performance (2019)



Portfolio Daily Performance (2020)



Portfolio Daily Performance (2021)



Portfolio Daily Performance (2022)



Indeed, the performance of the Portfolio Forecast was truly remarkable and surpassed all other comparable options. With an impressive Annual Return of 17.09%, it outperformed the second-best option by 1.02%. While this difference may seem small on its own, when scaled up to a million dollars, it translates into an additional value of \$10,000 per year, showcasing the substantial advantage of choosing the Portfolio Forecast.

Moreover, the Portfolio Forecast exhibited exceptional stability with a daily volatility of only 1.06%, in stark contrast to the second-best return option's higher volatility of 1.87%. This lower volatility indicates a smoother and more predictable investment journey, adding another layer of confidence in the superiority of the Portfolio Forecast's performance.

In summary, the significant Annual Return and lower Daily Volatility of the Portfolio Forecast make it an exceptional choice, offering potentially greater returns and reduced risk compared to other options considered during the analysis.

Parameter	Portfolio_Forecast
Annual Return	17.09%
Daily Volatility	1.06%

Parameter	Portfolio_Historical
Annual Return	14.35%
Daily Volatility	1.21%

Parameter	Max_Stock Forecast
Annual Return	15.95%
Daily Volatility	1.54%

Parameter	Max_stock Historical
Annual Return	16.07%
Daily Volatility	1.87%

Parameter	S&P500
Annual Return	11.28%
Daily Volatility	1.10%