Portfolio Selection Project Report

Daniel Dahan:345123624
Simon Bellilty: 345233563

Background

The objective of this project was to develop and implement a portfolio selection strategy for a given dataset of stock prices. The project is to create a portfolio strategy based on machine learning techniques, using historical stock data for the past five years to train the model.
Our proposed portfolio selection strategy employs machine learning models, including decision tree regression and linear regression with normalization and standardization. These models predict stock price changes based on historical data. They are relevant because they enable data-driven allocation of assets within the portfolio, considering both expected returns and risk. The strategy aims to optimize risk-adjusted returns, as measured by the Sharpe Ratio, and minimize portfolio volatility, as indicated by portfolio variance.

Data

The dataset employed in this project consisted of historical stock price data of the S&P 500 companies. This dataset encompassed five years of historical data, which was utilized as the training dataset for our portfolio selection strategy.
The test dataset, on the other hand, was meticulously separated to represent a specific evaluation period. In this case, the test data comprised stock data for the 22 days following the five-year training period. This division allowed us to evaluate the performance of our strategy on a distinct and limited dataset, ensuring that our approach could adapt to unforeseen market conditions and deliver reliable results in a real-world scenario.

Methods

Our portfolio selection strategy involved the following key steps:

1. **Data Preprocessing:** We downloaded historical stock data for a selection of S&P 500 companies using Yahoo Finance. This data included daily stock prices, trading volumes, open and close prices, and high and low prices. We processed this data to calculate additional features such as daily price changes and percentage changes.

2. **Model Training:** In the training phase, we developed a machine learning model for each individual stock in the dataset. We used a decision tree regression model and linear regression models with both normalization and standardization preprocessing steps. The goal was to predict the stock's price changes based on its historical features.

3. **Model Evaluation:** We evaluated the performance of each model using metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). RMSE was chosen as the primary evaluation metric for its ability to measure the accuracy of the price change predictions. Then we preserved the top-performing model for each individual stock individually.

4. **Portfolio Allocation:** In the portfolio allocation phase, we calculated the expected returns and risk for each individual stock based on the model predictions and historical data. We

then assigned weights to each stock in the portfolio based on a combination of predicted returns and risk.

5. **Normalization of Weights:** The weights assigned to each stock were normalized to ensure they summed to 1, representing a percentage-based allocation.

6. **Sharpe Ratio Calculation:** Finally, we calculated the Sharpe Ratio of the portfolio using the expected returns, portfolio variance, and a risk-free rate (assumed to be 0 for simplicity).

Experiments and Results

During the project, we conducted several experiments to fine-tune our portfolio selection strategy. These experiments included training various machine learning models for each stock, evaluating model performance, and adjusting portfolio allocation weights. The primary metrics used for evaluation were RMSE and the Sharpe Ratio.

In the evaluation on the test dataset, we achieved:

```
Sharp Ratio:  0.3773110240884747
Portfolio Variance:  0.0002135805610291199
```

The Sharpe Ratio holds a pivotal role in finance, shedding light on the risk-adjusted returns of an investment portfolio. A Sharpe Ratio of 0.3773 signifies that, on average, our portfolio generated returns that outperformed the risk-free rate by this factor while accounting for portfolio risk. This suggests that our strategy delivered positive returns relative to the level of risk assumed during the evaluation period.
Portfolio variance serves as a gauge for portfolio volatility, with lower values indicating reduced risk. In our case, the portfolio variance of 0.0002136 reflects a portfolio that exhibited relatively low volatility during the evaluation period, a noteworthy attribute of our strategy.

These findings signify that our portfolio selection strategy effectively generated risk-adjusted positive returns during the test period, as evidenced by the Sharpe Ratio. Furthermore, our portfolio demonstrated stability, as indicated by its low portfolio variance.
The results of our experiments demonstrated that our portfolio selection strategy successfully allocated assets based on the expected returns and risk associated with each stock. The allocation weights were adjusted to ensure a well-diversified portfolio. The calculated Sharpe Ratio served as a measure of risk-adjusted returns for our portfolio.

Conclusion

In conclusion, our strategy relied on individual models for each stock, taking into account historical data and predicted price changes. The portfolio allocation was based on a combination of expected returns and risk, resulting in a well-balanced and diversified portfolio.
With a positive Sharpe Ratio and a portfolio characterized by low volatility, our strategy showcased its ability to generate risk-adjusted returns.