QF2103 Group 14

Charmi Nagar, Chen Hung-Yu, Choo Qi En Jonathan, Huynh Minh Thuan, Jax Lee Le Sheng

April 10, 2024

Abstract

This study explores two distinct trading strategies in financial markets: day trading and swing trading. The first trading strategy is to predict signals solely based price movements using different models and strategies based on price movements. In contrast, the second trading strategy relies on forecasting methods to capture medium-term price trends over several days to weeks. Both strategies aim to capitalize on market inefficiencies but differ in their time horizons and execution methodologies.

1 Team Roles and Responsibilities

	Roles and Responsibilities
Charmi Nagar	Backtesting, Model solutions, Analysis and solutions of model shortfalls and risks
Chen Hung-Yu	Collected and preprocessed additional data features
	Implemented, backtested, and optimized GB Classifier Model and DCA
Choo Qi En Jonathan	Implemented, backtested and optimised Fama Fench Model
Huynh Minh Thuan	
Jax Lee Le Sheng	Classification/Regression/Time Series Models.
	Analysis of indicators. Day/Swing Trading Strategies. Backtesting

2 Technical Indicator

2.1 Slow Stochastic Oscillator (SlowK and SlowD):

Measures the momentum of a financial instrument's price in relation to its recent trading range over a specified period.

$$SlowK = 100 \times \left(\frac{\text{Close} - \text{LowestLow}}{\text{HighestHigh} - \text{LowestLow}} \right)$$

$$SlowD = \text{Moving average of SlowK over a specified period}$$

2.2 KDJ (K-D):

A variation of the Stochastic Oscillator that incorporates a smoothing factor.

$$K = \text{Smoothing factor} \times \left(\frac{\text{Current Close} - \text{Lowest Low}}{\text{Highest High} - \text{Lowest Low}} \right)$$

$$D = \text{Moving average of K over a specified period}$$

2.3 Bollinger Bands (Upper Band, Middle Band, Lower Band, Band Width):

A volatility indicator that consists of a middle band (simple moving average) with upper and lower bands representing standard deviations away from the middle band.

$$\text{Upper Band} = \text{Middle Band} + (\text{Standard Deviation} \times K)$$

$$\text{Lower Band} = \text{Middle Band} - (\text{Standard Deviation} \times K)$$

$$\text{Band Width} = \frac{\text{Upper Band} - \text{Lower Band}}{\text{Middle Band}}$$

2.4 BIAS (20-day, 60-day, 5-day):

Measures the percentage difference between a financial instrument's current price and its moving average over a specified period.

$$BIAS20 = \left(\frac{\text{Close} - \text{MA20}}{\text{MA20}}\right) \times 100$$

$$BIAS60 = \left(\frac{\text{Close} - \text{MA60}}{\text{MA60}}\right) \times 100$$

$$BIAS5 = \left(\frac{\text{Close} - \text{MA5}}{\text{MA5}}\right) \times 100$$

2.5 On-Balance Volume (OBV):

Measures buying and selling pressure by adding volume on up days and subtracting volume on down days.

$$OBV = \text{Previous OBV} + \text{Volume (if Close} > \text{Previous Close)} - \text{Volume (if Close} < \text{Previous Close)}$$

2.6 Moving Average Convergence Divergence (MACD):

Compares two moving averages to signal potential buy and sell opportunities.

$$\text{MACD Line} = \text{12-day EMA} - \text{26-day EMA}$$

2.7 MACD Signal Line:

Moving average of the MACD line over a specified period.

$$\text{MACD Signal Line} = \text{Moving average of MACD Line over a specified period}$$

2.8 MACD Histogram:

Represents the difference between the MACD Line and the MACD Signal Line.

$$\text{MACD Histogram} = \text{MACD Line} - \text{MACD Signal Line}$$

2.9 Relative Strength Index (RSI):

Measures the magnitude of recent price changes to evaluate overbought or oversold conditions.

$$RSI = 100 - \left[\frac{100}{1 + \left(\frac{\text{Average Gain}}{\text{Average Loss}}\right)}\right]$$

3 Trading Strategy 1: Predicting Price movements

Our first strategy solely focuses on generating signals of price movements and using it to formulate a trading strategy. We discuss our efforts on using different price movement strategies on different machine learning models.

3.1 Train Test Split

The dataset is divided into training and testing using 80/20 as required, while preserving the time order.

3.2 Initial Model

The initial trading strategy has features selected based on the Fama-French 3 factor model.

3.2.1 Fama-French 3 Factor Model

The Fama-French 3 factor model is an asset pricing model developed by Eugene Fama and Kenneth French and it extends the Capital Asset Pricing Model by including risk and performance factors.

The equation of the model is as follows:

$$r = r_f + \beta_1(r_m - r_f) + \beta_2(SMB) + \beta_3(HML) + \epsilon$$

Where

- r is the expected rate of return
- r_f is the risk-free rate
- β are the factor's coefficients
- $r_m - r_f$ is the market risk premium
- SMB is the historic excess returns of small-cap companies over large-cap companies
- HML is the historic excess returns of value stocks over growth stocks

3.2.2 Trading Strategy

Using the Fama-French 3 Factor model and data on the 3 factors, the expected returns of a stock was predicted using OLS by fitting the 3 factors.

The simple strategy is as follows: if the predicted returns > actual returns, buy the stock. Otherwise, sell.

3.2.3 Data Preparation

Data regarding the 3 factors were collected in `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html` under the "Fama/French 3 Factors [Daily]" section.

Due to parsing errors, the data collected was manually edited so that it could be worked on.

3.2.4 Model Training

Ordinary Least Squares: The beta coefficients were estimated by regressing the percentage change of returns on the 3 factors using the training data. Percentage change of returns in the testing data was then predicted. Signal were generated based on the strategy above. If predicted returns > actual returns, a 1 is generated, otherwise -1 is generated.

3.2.5 Model Evaluation

Backtesting Results

Using the signals generated, the strategy was backtested against the benchmark strategy of just buying and holding the stock. The picture below shows the backtesting results.

```
Backtest results for returns
Returns AAPL.O           1.574734
Strat Returns AAPL.O     0.743315
dtype: float64

Returns MSFT.O           1.723348
Strat Returns MSFT.O     1.488425
dtype: float64

Returns INTC.O           1.333065
Strat Returns INTC.O     1.261794
dtype: float64

Returns AMZN.O           2.090903
Strat Returns AMZN.O     1.085851
dtype: float64

Returns GS.N           1.305148
Strat Returns GS.N     1.168354
dtype: float64
```

(a) Returns

```
Volatility results
Returns AAPL.O           0.012349
Strat Returns AAPL.O     0.012390
dtype: float64

Returns MSFT.O           0.012188
Strat Returns MSFT.O     0.012228
dtype: float64

Returns INTC.O           0.015224
Strat Returns INTC.O     0.015236
dtype: float64

Returns AMZN.O           0.015044
Strat Returns AMZN.O     0.015156
dtype: float64

Returns GS.N           0.013873
Strat Returns GS.N     0.013860
dtype: float64
```

(b) Volatility

Figure 1: Strategy Performance

As shown in the results, the strategy made profit and losses depending on the stock, but none of them outperform the benchmark strategy in both returns and volatiltiy.

Model Performance

The performance of the model was generated using statsmodel.api's summary function on the testing data. An example below is shown for apple stock.

```
Model summary for AAPL.O

                            OLS Regression Results
==============================================================================
Dep. Variable:         Returns AAPL.O   R-squared:                       0.427
Model:                            OLS   Adj. R-squared:                  0.423
Method:                 Least Squares   F-statistic:                     105.3
Date:                Wed, 10 Apr 2024   Prob (F-statistic):           6.04e-51
Time:                        19:50:30   Log-Likelihood:                 1393.0
No. Observations:                 428   AIC:                            -2778.
Df Residuals:                     424   BIC:                            -2762.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0003      0.000      0.578      0.564      -0.001       0.001
Mkt-RF         0.0109      0.001     16.332      0.000       0.010       0.012
SMB           -0.0021      0.001     -2.355      0.019      -0.004      -0.000
HML           -0.0066      0.001     -7.280      0.000      -0.008      -0.005
==============================================================================
Omnibus:                      130.961   Durbin-Watson:                   1.849
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              845.404
Skew:                           1.143   Prob(JB):                   2.65e-184
Kurtosis:                       9.495   Cond. No.                         2.07
==============================================================================
```

Figure 2: Model Summary for Apple Stock

The R^2 value is low, suggesting that little variance in excess returns is explained by the 3 factors.

3.2.6 Optimisation Efforts Within Fama Strategy

Two attempts were made to generate better results. The first is to change the strategy by doing the opposite of the proposed strategy: buy when predicted returns < actual returns, sell when predicted returns > actual returns. The second is to include two more features by using the Fama French 5 factor model. Overall, the results for these two optimisation efforts follows roughly the same pattern as the non-optimised one.

3.3 Final Module

3.3.1 Data Preparation

1. Raw historical market data, including open, high, low, close, and volume, is collected from Yahoo Finance via YF API.
2. Stock price is adjusted for stock splits.
3. Split by individual stock
4. Generate target label 0 if next day close low and 1 if the next day close high.
5. Generate technical indicators by TA-lib.
6. Preprocess the signals and generate additional features. i.e. Take upper BBAND − middle BBAND

3.3.2 Model Training

Boosting Model (HistGradientBoostingClassifier): The transformed features are fed into a gradient boosting classifier to predict the next day's price movement. Boosting iteratively improves the model by fitting each weak learner to the residuals of the previous learner. The Hist variant is chosen for its ability to handle missing values and fast training speed. One model per stock was generated.

3.3.3 Model Evaluation

Once trained, the model is evaluated on the testing set to assess their performance in predicting the next day's price movement. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to measure model effectiveness.

3.3.4 Trading Strategy

Ensemble methods can not only provide prediction but can also output the predicted probability of classes. Inspired by this, we propose two strategies:
Naive

- Open long position at market open and close at market close if predicted 1
- Open short position at market open and close at market close if predicted 0

Long-Short

- Open long position in proportion to the 1 label probability
- Open short position in proportion to the 0 label probability
- Close both positions at market close

It is assumed that there are no transaction costs and stocks can be freely shorted.

Stock	Naive	Long-Short
AAPL	-0.21	-0.08
MSFT	-0.24	-0.07
INTC	-0.11	-0.01
AMZN	-0.15	-0.02
GS.N	-0.46	-0.29

Table 1: Log Returns

Stock	Naive	Long-Short
AAPL	0.01000	0.00124
MSFT	0.01052	0.00162
INTC	0.01249	0.00462
AMZN	0.01277	0.00018
GS.N	0.01208	0.00459

Table 2: Volatility in daily log returns

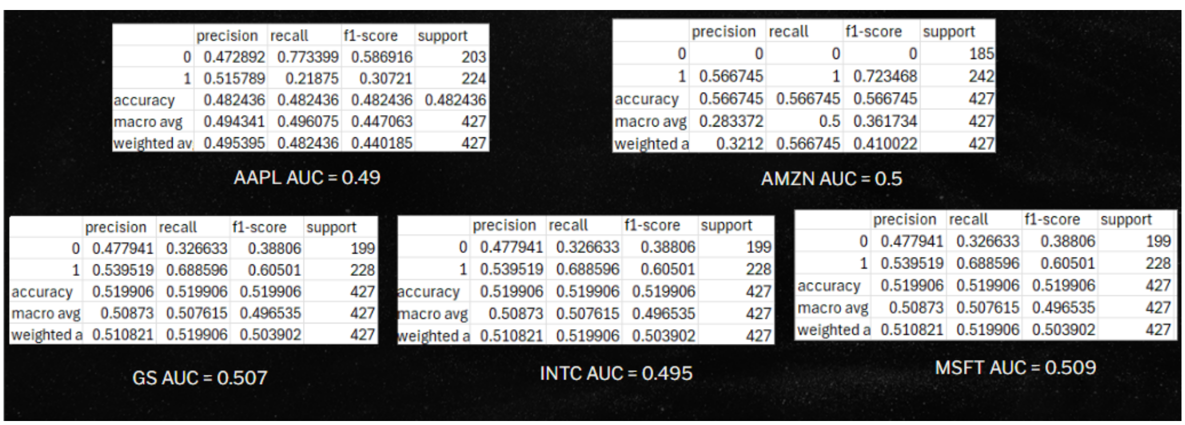
3.3.5 Model Evaluation

Backtesting Results

The log returns and volatility of daily returns are calculated and displayed in Table 1 and Table 2 for both trading strategies. It is observed that both strategies generate negative returns overall. Long-short, as expected, has a lower negative return and volatility.

Model Performance

The performance of the models is generated using "sklearn.metrics.classification_report." The models all have an AUC of around 0.5, indicating that their predictions are close to completely random. This result is expected because of the random walk nature of the stock market and the small training set with daily data.



	precision	recall	f1-score	support
0	0.472892	0.773399	0.586916	203
1	0.515789	0.21875	0.30721	224
accuracy	0.482436	0.482436	0.482436	0.482436
macro avg	0.494341	0.496075	0.447063	427
weighted av	0.495395	0.482436	0.440185	427

AAPL AUC = 0.49

	precision	recall	f1-score	support
0	0	0	0	185
1	0.566745	1	0.723468	242
accuracy	0.566745	0.566745	0.566745	427
macro avg	0.283372	0.5	0.361734	427
weighted a	0.3212	0.566745	0.410022	427

AMZN AUC = 0.5

	precision	recall	f1-score	support
0	0.477941	0.326633	0.38806	199
1	0.539519	0.688596	0.60501	228
accuracy	0.519906	0.519906	0.519906	427
macro avg	0.50873	0.507615	0.496535	427
weighted a	0.510821	0.519906	0.503902	427

GS AUC = 0.507

	precision	recall	f1-score	support
0	0.477941	0.326633	0.38806	199
1	0.539519	0.688596	0.60501	228
accuracy	0.519906	0.519906	0.519906	427
macro avg	0.50873	0.507615	0.496535	427
weighted a	0.510821	0.519906	0.503902	427

INTC AUC = 0.495

	precision	recall	f1-score	support
0	0.477941	0.326633	0.38806	199
1	0.539519	0.688596	0.60501	228
accuracy	0.519906	0.519906	0.519906	427
macro avg	0.50873	0.507615	0.496535	427
weighted a	0.510821	0.519906	0.503902	427

MSFT AUC = 0.509

Figure 3: Model Performances

3.3.6 Optimization Efforts

I noticed that the models have a high accuracy score on the training set, which indicates overfitting. I adjusted the relevant parameters of the model by increasing L2 regularization, increasing min samples, and decreasing max depth. In addition, I observed that the log-returns of the model are not completely random. Instead, they all consistently generate negative returns across the five stocks. This means if we invert our current strategy, we can achieve positive returns. This technique may seem counterintuitive at first, but we can rationalize it as our model makes incorrect predictions more often when there are large movements in stock price and more predictions correct when the stock price movement is more stable. Therefore, we can leverage the spread to generate positive returns. The **Inverted Naive Strategy** is compared with **Dollar Cost Averaging (DCA)** and **Buy-and-Hold**.

Stock	Inverted	Buy-and-Hold	DCA
AAPL	0.21	0.45	0.21
MSFT	0.24	0.53	0.27
INTC	0.11	0.32	0.21
AMZN	0.15	0.72	0.50
GS.N	0.46	0.24	-0.05

Table 3: Log Returns for Inverted Naive Strategy

Results show that the Inverted Naive Strategy can produce log returns on par with DCA and sometimes even outperform. It also boasts a significant advantage over the other two strategies, which is low volatility. This strategy enters and exits the market daily so that it can avoid periods of time with a large unrealized loss that can happen in the other two strategies.

3.3.7 Final Strategy

Inverted naive (single-side) strategy

3.3.8 Limitations

- Hard to predict the next day closing prices: only have data from the previous day, and stock prices can be highly affected by events that happen on the same day
- Bias in stocks chosen: All stocks listed in the US, in the tech industry, well-known blue chip stocks

These are some limitations that are inherent to the dataset or model.

4 Trading Strategy 2: Swing Trading Via Forecasting

Swing trading involves exploiting short- to medium-term price fluctuations in financial markets. Traditional forecasting approaches rely heavily on mathematical models such as ARIMA and ETS for forecasting future price movements. However, these models often struggle with highly volatile and noisy datasets, In response to these challenges, we investigates an alternative strategy.

4.1 Decision Making: Long or Short Position Selection

In swing trading, the decision to take a long or short position is crucial and heavily influences the profitability of each trade. The decision to take a long or short position hinges on the duration with the highest forecasted returns. If the analysis indicates that the maximum forecasted returns occur over a positive price trajectory within the trading horizon, a long position is initiated. Conversely, if the highest forecasted returns coincide with a downward price trajectory, a short position is taken to capitalize on potential price declines.

4.2 Initial Forecasting Models

Despite the importance of accurate price forecasting in swing trading, traditional time series models such as ARIMA, ETS, and LSTM have shown limitations in effectively capturing the dynamics of the market.

4.2.1 ARIMA (AutoRegressive Intergrated Moving Average) Model

ARIMA is a widely-used time series forecasting technique that combines autoregressive and moving average components. Despite its popularity, the ARIMA model struggled to extract relevant patterns from the dataset utilized in this study. The inherent noise-to-signal ratio in the data made it challenging for the model to discern actionable insights, resulting in suboptimal forecasting performance.

4.2.2 ETS (Error-Trend-Seasonality) Model

ETS models are another class of traditional forecasting methods commonly employed in time series analysis. These models decompose time series data into error, trend, and seasonality components to forecast future values. However, similar to ARIMA, the ETS model failed to adequately capture the underlying patterns in the dataset. The lack of discernible trends or seasonality posed significant challenges in generating accurate forecasts for swing trading decisions.

4.2.3 LSTM (Long Short-Term Memory) Model

LSTM, a type of recurrent neural network (RNN), is known for its ability to capture long-term dependencies in sequential data. Despite its complexity and potential for capturing nonlinear relationships, the LSTM model encountered similar difficulties in extracting meaningful insights from the dataset, even with the inclusion of additional signals as discussed in Chapter 2. In an attempt to mitigate this issue, an alternative approach was explored, involving the discretization of log returns through binning to tokenizing them akin to methods used in Natural Language Processing (NLP). While the tokenization of log returns offered a novel perspective and facilitated the incorporation of sequential data into the LSTM model, the high noise-to-signal ratio and lack of discernible statistical properties in the dataset persisted as significant obstacles. Consequently, the LSTM model continued to struggle in extracting actionable insights and generating accurate forecasts, despite the innovative preprocessing technique.

4.3 Final Model

Amidst the challenges encountered with conventional forecasting models such as ARIMA, ETS, and LSTM, a novel approach known as forecasting by analogy emerged as the cornerstone of the final trading strategy. This section delves into the concept of forecasting by analogy and its application in overcoming the limitations posed by traditional modeling techniques.

4.3.1 Concept of Forecasting by Analogy

Forecasting by analogy is a heuristic approach that draws parallels between current market conditions and historical precedents to make informed predictions about future price movements. Unlike conventional time series models that rely solely on mathematical algorithms, forecasting by analogy leverages qualitative analysis, historical data examination, and pattern recognition techniques to identify recurring patterns and trends in the market.

4.3.2 Advantages and Implication

- Flexibility: The qualitative nature of forecasting by analogy allows for adaptability to changing market conditions and unforeseen events.
- Forecasting by analogy acknowledges the inherent uncertainties and complexities of financial markets, embracing a pragmatic approach to decision-making.

4.3.3 Model Training

No training was required. The following is a brief description of the method as mention in its paper.

The method is based on identifying a historical window in one of the other time series that is highly correlated with the target time series, just before the forecast horizon. Let the size of this window be M.

Here is the formal description of the method:

1. Let $Y' = (y_{m-M+1}, \ldots, y_m)$ be the last M observations of Y.
2. For every series $X = (x_1, \ldots, x_m)$:
   a) For i from 1 to $m - h - M + 1$, compute the correlation coefficient between the window $(x_i, \ldots, x_{i+M-1})$ and Y'.
3. Let Z be the series containing the window most correlated with Y'. Let $Z' = (z_j, \ldots, z_{j+M-1})$ be that window.
   a) Let $Z'' = (z_{j+M}, \ldots, z_{j+M+h-1})$ be the continuation of Z' with length h.
   b) Replace the forecast for Y with

$$(Z'' - \bar{Z}')\frac{\sigma(Y')}{\sigma(Z')} + \bar{Y}'$$

   Here is *an example* to visually depict the above. Note that Vietnam may not be the best for Indonesia, and neither is the window indicated optimal; it's just an example.

Figure 4: Forecasting by Analogy Method

4.4 Optimisation Strategies

Forecasting by analogy may overlook the temporal evolution of data and fail to capture evolving patterns adequately. To address this limitation, we propose a method that assigns varying weights to historical data points based on their temporal proximity to the target forecast period.

$$\rho_w = \frac{\sum_{i=1}^{n} w_i(X_i - \bar{X}_w)(Y_i - \bar{Y}_w)}{\sqrt{\sum_{i=1}^{n} w_i(X_i - \bar{X}_w)^2}\sqrt{\sum_{i=1}^{n} w_i(Y_i - \bar{Y}_w)^2}} \tag{1}$$

4.5 Final Results

ticker	Percentage Profit in %
aapl	0.78
msft	0.01
intc	0.10
amzn	-0.08
gs	0.20

Table 4: Forecasting by Analogy for Swing Trading

4.6 Limitations

Despite our efforts in developing a forecasting by analogy method, several limitations emerged during our experimentation:

4.6.1 Lack of Stop Loss Function

One significant limitation is the absence of a stop-loss mechanism in our approach. This omission resulted in situations where the predicted direction of the stock deviated substantially from the actual direction, leading to significant losses. Incorporating a stop-loss function could potentially mitigate these losses by automatically terminating trades when deviations exceed predefined thresholds.

4.6.2 Model Training

Our method's reliance on a restricted set of historical datasets posed another limitation. We primarily utilized a predefined set of stocks for forecasting, which may not have adequately captured the complexity and dynamics of the market. The limited scope of our dataset could have introduced biases and undermined the reliability of our forecasts. Additionally, our forecasting methodology may have been susceptible to overfitting. By selecting historical analogs based solely on their correlation coefficients, we may have inadvertently overfit the model to past data, resulting in poor performance on unseen data. Robust cross-validation techniques and regularization methods should be explored to mitigate overfitting and improve the generalization of our forecasting model.

4.7 Future Directions

Developing and integrating more sophisticated risk management strategies, such as dynamic stop-loss mechanisms or position sizing algorithms, could enhance the robustness of our forecasting framework. By incorporating adaptive risk controls, we can better manage downside risk and improve the overall performance of our trading strategy.

5 Potential Challenges and Risk of Trading Strategy

In the formulation and execution of trading strategies, especially those dependent on quantitative analyses such as day trading and swing trading, a thorough understanding and management of inherent challenges and risks are paramount for ensuring the strategies' robustness and long-term viability. This section delves into critical aspects related to market dynamics, the efficacy of models, and the regulatory framework, each of which holds significant implications for the outcomes of trading activities.

- ***Market Volatility and Liquidity Risks***

Market volatility represents one of the foremost unpredictable factors in trading, offering both opportunities for gain and significant exposure to risk. Particularly, day traders, who leverage short-term price movements for profit, may encounter increased volatility leading to slippage. This term refers to the discrepancy between a trade's anticipated execution price and the actual price achieved. Such variance can diminish potential profits or amplify losses, highlighting the precarious nature of relying on short-term market movements.

Equally, liquidity risk constitutes a major concern, especially pertinent to swing trading strategies that maintain positions over extended durations. In scenarios where markets or specific instruments exhibit lower liquidity, the challenge of entering or exiting positions at preferred price points escalates. This can result in trade delays or the necessity to settle for less favorable prices. Both outcomes can adversely influence the performance of trading strategies, underscoring the critical need for strategic planning and risk management in the face of market volatility and liquidity constraints.

- ***Model shortfall***

When crafting trading strategies that lean on machine learning and statistical models, there is a need to navigate to a delicate balance. There's a fine line between creating models that are either too intricate or too simplistic. On one end, we have models that become so detailed they start seeing patterns in what is essentially random noise, focused on historical data but failing when they encounter new scenarios. On the other, some models may barely scratch the surface of the market's complexities, resulting in lackluster forecasts both for past and future events. Sometimes, the results are just random. Financial markets are inherently chaotic systems, influenced by countless variables and subject to unexpected events. In this context, a model's failure to predict future movements accurately isn't always a sign of its inadequacy in capturing market dynamics. Instead, it may simply reflect the unpredictable nature of the markets themselves.

6 Conclusion

We have delved into the exploration of two prominent trading strategies within the financial markets, using forecasting to exploit market inefficiencies. Despite encountering challenges like model overfitting and market volatility, these hurdles mainly guided our project towards better strategies and risk management.