Predictive Hedging Strategies Using LSTM for Automated Trading

Authors: Sanmitra Sur, Amit Bera, Subham Roy

Supervisor: Dr. Gopal Krishna Basak

Date: December 16, 2024

Abstract

This study explores the development of a predictive hedging framework by leveraging time-series forecasting models and trading signal generation techniques. A Long Short-Term Memory (LSTM) model was implemented to forecast the *Close Price* of financial assets using a set of selected features derived through rigorous Exploratory Data Analysis (EDA) and feature selection methods. Hyperparameter tuning was performed to optimize the model's performance, achieving a Root Mean Squared Error (RMSE) of 0.1097, a Mean Absolute Error (MAE) of 0.0804, and a Mean Absolute Percentage Error (MAPE) of 11.66%. Additionally, a systematic trading signal generation strategy was designed based on predicted price movements to produce actionable BUY, SELL, and HOLD signals. This framework demonstrates the applicability of LSTM-based models for predictive trading and automated hedging strategies in dynamic financial markets. Future work proposes further enhancements, including the incorporation of external macroeconomic indicators, ensemble models, and blockchain-based smart contracts for automated execution.

1 Introduction

In financial markets, price volatility and uncertainty pose significant challenges to both individual and institutional investors. Hedging strategies, which aim to mitigate financial risks, require accurate predictions of asset prices and timely execution of trading decisions. Traditional methods for price forecasting often rely on statistical models such as ARIMA and GARCH. While effective for capturing linear trends, these methods often fail to model complex, non-linear dependencies in time-series data. Recent advancements in deep learning, particularly the development of Long Short-Term Memory (LSTM) networks, have demonstrated superior performance in capturing temporal patterns and predicting future asset prices.

This study focuses on building a predictive hedging framework by combining LSTM-based time-series forecasting with a systematic trading signal generation strategy. Key objectives include:

- Accurately predicting the *Close Price* of financial assets using LSTM networks.
- Generating actionable BUY, SELL, and HOLD trading signals based on model predictions.
- Evaluating the model's performance through appropriate metrics and validating its applicability to real-world financial markets.

The framework is developed in a series of well-defined stages:

1. **Exploratory Data Analysis (EDA):** Insights are extracted from price movements and technical indicators such as Bollinger Bands, RSI, and MACD to understand trends, volatility, and correlations.
2. **Feature Selection:** Features contributing significantly to the prediction of *Close Price* are selected using Correlation Analysis, Recursive Feature Elimination (RFE), and Mutual Information methods.
3. **LSTM Model Development:** An LSTM-based model is implemented and optimized using hyperparameter tuning to minimize prediction errors.
4. **Trading Signal Generation:** Trading signals (BUY, SELL, HOLD) are generated using pre-defined thresholds based on the predicted price changes.

The results demonstrate that the LSTM model effectively predicts price movements, achieving low error rates and generating reliable trading signals. These findings underscore the potential of deep learning models for predictive trading and automated hedging strategies.

The remainder of this report is organized as follows: Section 3 presents the EDA and insights from technical indicators. Section 4 describes the feature selection process. Section 6 details the development and optimization of the LSTM model. Section 7 discusses the trading signal generation framework. Finally, Section 8 summarizes the key findings, contributions, and proposed future work.

2 Data Preprocessing

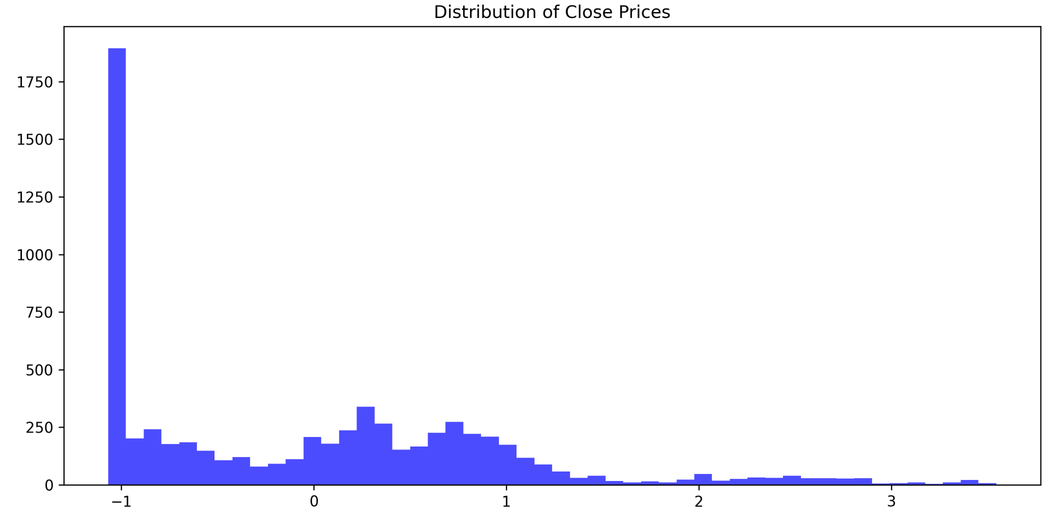
The dataset used in this study contains daily stock data for **SBI (State Bank of India)**, with fields including *Open*, *High*, *Low*, *Close*, and *Volume*. Preprocessing steps were performed to ensure the data was clean, consistent, and suitable for time-series forecasting. The key preprocessing steps are outlined as follows:

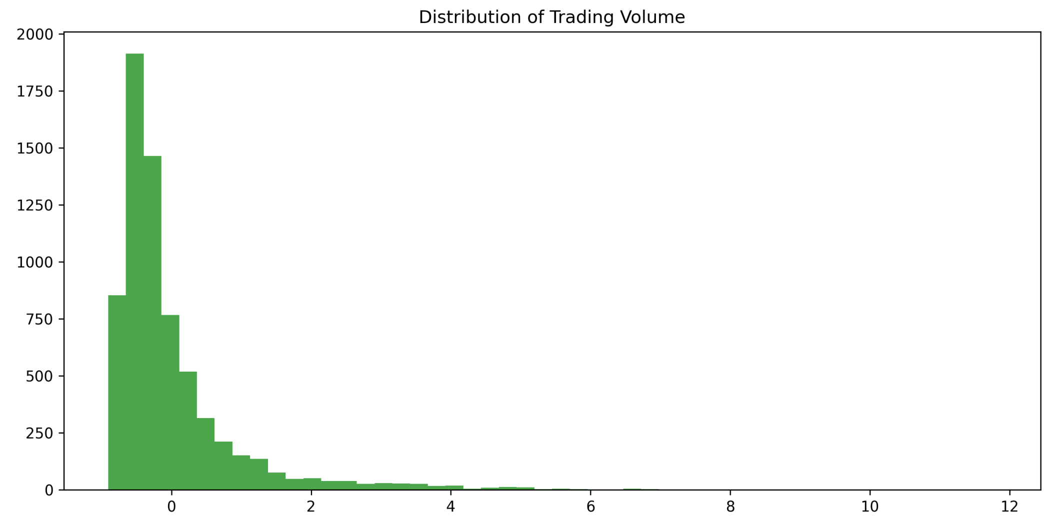
1. **Date Handling:** The *Date* column was converted to a datetime type and set as the index to facilitate time-series analysis.
2. **Handling Missing Values:** Missing values were handled using the following strategies:
   - Forward fill (`ffill`) was applied to propagate the last valid observation.
   - Backward fill (`bfill`) was used to fill any remaining missing values.
3. **Duplicate Removal:** Duplicate records in the dataset were identified and removed to avoid redundancy.
4. **Feature Standardization:** Numerical features (*Open*, *High*, *Low*, *Close*, *Volume*) were standardized using the **StandardScaler** to ensure a mean of 0 and standard deviation of 1. This step ensures all features are on a similar scale, which is critical for improving model performance.
5. **Feature Engineering:** Additional features were engineered to provide temporal context and improve the predictive capability of the model:
   - **Rolling Statistics:** A 7-day rolling mean and standard deviation of the *Close Price* were computed to capture short-term trends and volatility:
     - *rolling_mean_close*: 7-day rolling average of *Close Price*.
     - *rolling_std_close*: 7-day rolling standard deviation of *Close Price*.
   - **Lagged Feature:** A 1-day lag of the *Close Price* (*lag1_close*) was added to provide historical price information as a predictive feature.
   - **Day of Week:** A categorical feature representing the day of the week (0 = Monday, 6 = Sunday) was added to capture weekly patterns in the data.
6. **Saving the Preprocessed Data:** The cleaned and transformed dataset was saved to a new CSV file (`preprocessed_data.csv`) for use in further analysis and model development.

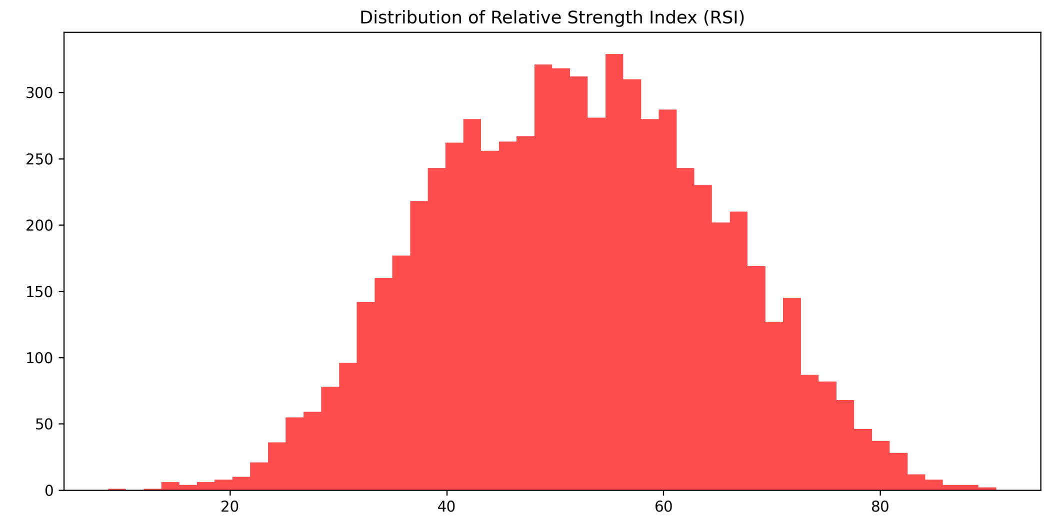
Conclusion: The preprocessing steps ensured that the dataset was free of missing values and duplicates, standardized for consistency, and enriched with engineered features such as rolling statistics, lagged values, and day-of-week indicators. These enhancements provide a strong foundation for accurate time-series forecasting and trading signal generation.

3 Exploratory Data Analysis

3.1 Distribution of Key Features







Observations and Significance:

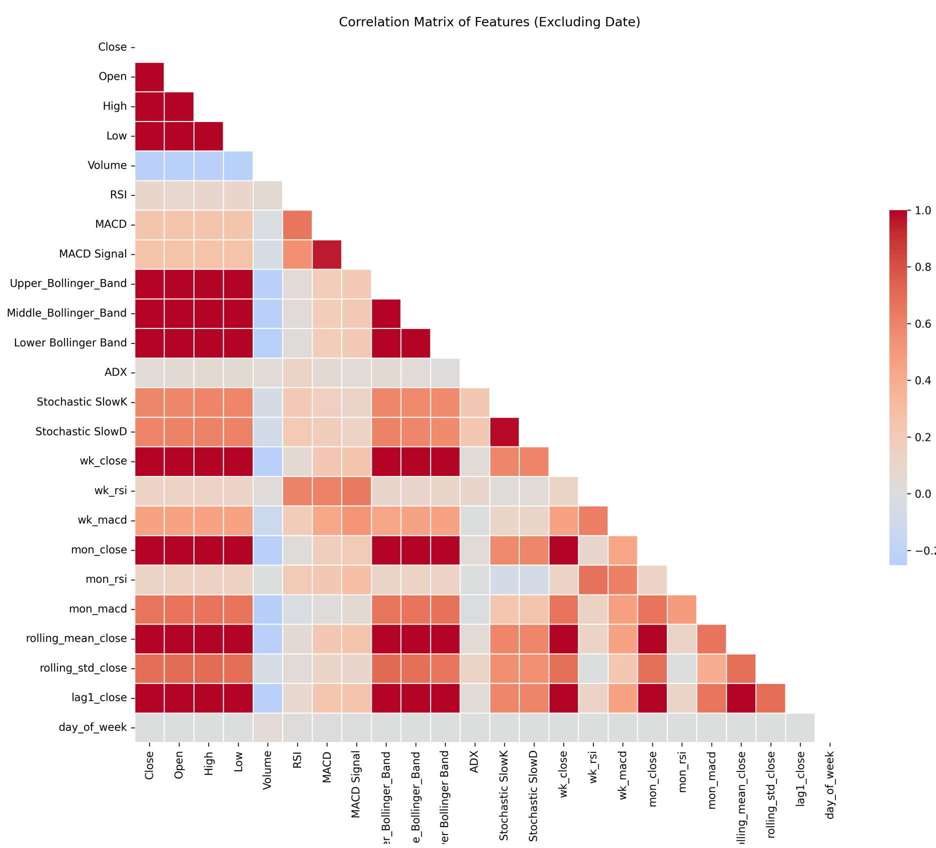
- **Close Prices:** The Close Prices distribution is significantly left-skewed, with the majority of values concentrated near -1 and a long tail extending to the right. *Significance:* This skewness may indicate a prevalence of smaller price fluctuations and a few extreme positive returns. Such a pattern is often seen in highly volatile or speculative markets where occasional large gains are observed alongside frequent minor movements.

- **Trading Volume:** The Trading Volume distribution is heavily right-skewed, with most observations clustered near zero and a few extreme spikes. *Significance:* This signifies that, on most days, trading activity is low, while sporadic periods of very high volumes occur. These spikes could correspond to news events, earnings announcements, or other factors driving significant market activity.

- **Relative Strength Index (RSI):** The RSI distribution shows a relatively symmetric bell-shaped curve, centered around mid-range values (40–60). *Significance:* A bell-shaped RSI indicates that the asset's price generally oscillates between overbought (above 70) and oversold (below 30) levels, spending most of the time in neutral territory. This suggests relatively stable price movements without prolonged periods of extreme market conditions.

Overall Interpretation: The distributions of these key features suggest a market environment where prices tend to fluctuate with occasional significant returns, low to moderate trading activity dominates, and relative stability prevails as indicated by the RSI values. These insights are critical for designing predictive models or hedging strategies that account for volatility and trading behavior.

3.2 Correlation Matrix of Features

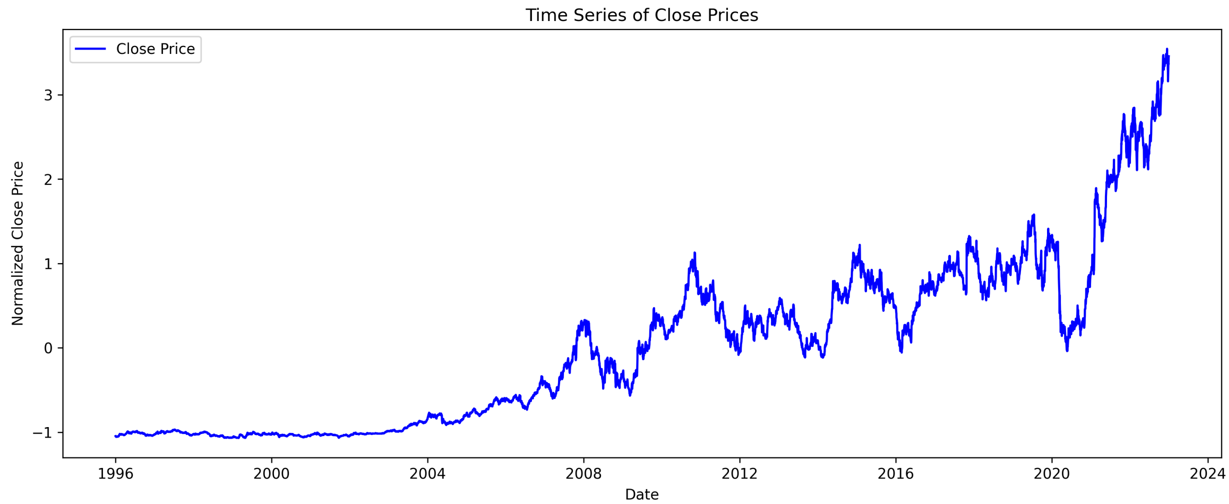


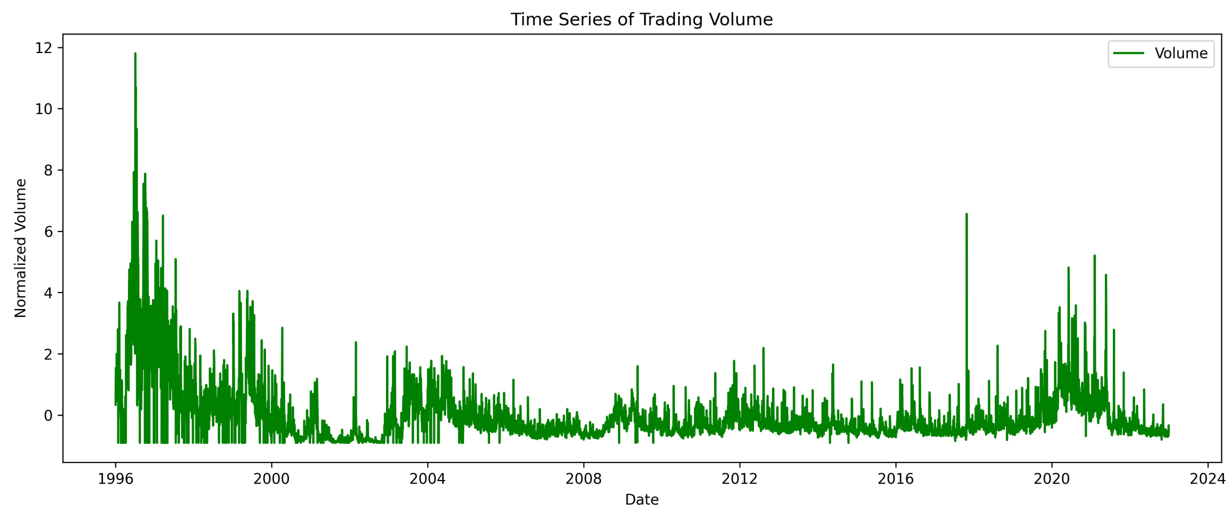
Observations and Significance:

- **Close, Open, High, and Low Prices:** These features exhibit a strong positive correlation (close to 1), which is expected since price metrics for a single asset tend to move together. *Significance:* This indicates that any predictive model can safely consider these features as redundant, avoiding multicollinearity issues by selecting only one or two.
- **Volume:** Volume shows weak or negative correlations with price-based metrics (e.g., Close, Open), suggesting that trading volume does not directly align with price changes. *Significance:* This may imply that changes in trading volume are more event-driven or indicative of liquidity rather than price trends.
- **MACD and Bollinger Bands:** MACD (Moving Average Convergence Divergence) and Bollinger Bands show high correlations with price metrics like Close and Open. *Significance:* These technical indicators are inherently derived from price, reaffirming their strong association and relevance for identifying trends.
- **Rolling Statistics (Mean and Std):** Rolling mean and rolling standard deviation of close prices are moderately correlated with current close prices. *Significance:* These rolling statistics highlight short-term trends and price volatility, which can be useful for time-series modeling.
- **Day of Week:** The feature ‘day_of_week‘ shows near-zero correlation with most other features. *Significance:* This suggests that day-based temporal effects might not have a strong linear relationship with prices or trading metrics, though non-linear patterns may still exist.
- **Inverse Correlations:** Certain features (e.g., RSI and Stochastic indicators) display mild negative correlations with price metrics, aligning with their purpose of identifying overbought or oversold conditions. *Significance:* These inverse correlations suggest RSI and Stochastic indicators can help capture turning points or trend reversals.

Overall Interpretation: The correlation matrix highlights key relationships between features, confirming strong dependencies among price-based metrics and derived indicators. Weak correlations, such as those involving volume and day-based features, emphasize the need for careful feature selection when building predictive models to ensure efficiency and avoid redundancy.

3.3 Time Series of Close Prices and Trading Volume





Observations and Significance:

- **Close Prices:** The Close Prices exhibit a long-term upward trend with noticeable fluctuations and periods of volatility, especially after 2008. *Significance:* This long-term trend suggests steady market growth, interspersed with corrections or crashes likely driven by macroeconomic events, such as the 2008 financial crisis, 2020 pandemic crash, and subsequent recoveries.

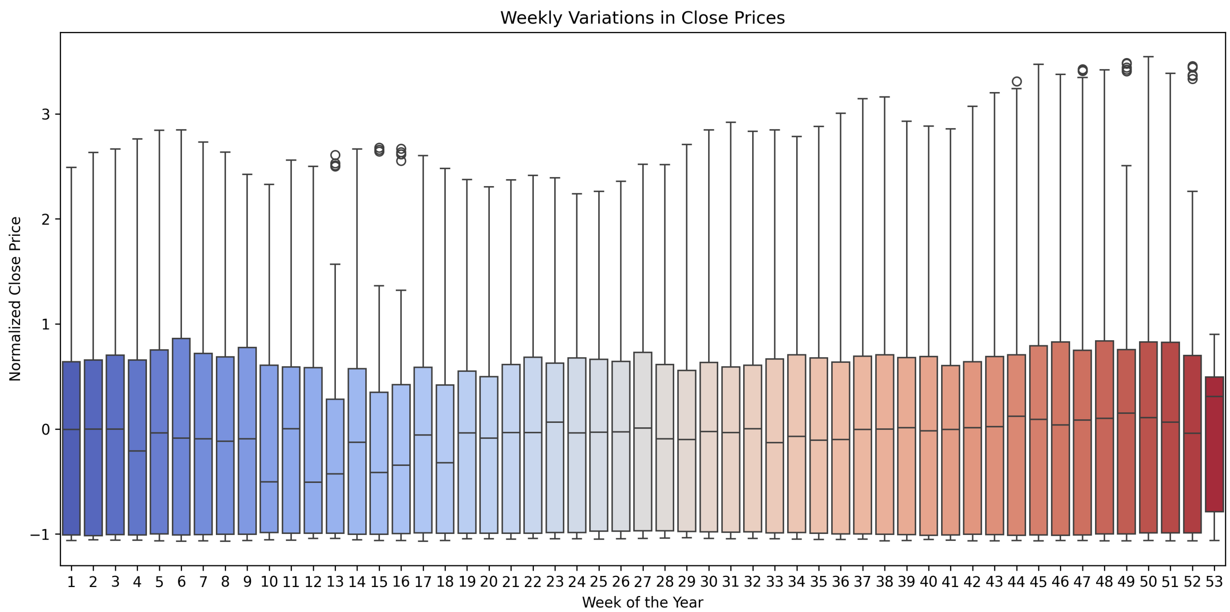
- **Volatility in Close Prices:** Close price volatility increases significantly after 2004, with larger peaks and troughs observed between 2008–2024. *Significance:* The increased volatility highlights changing market conditions, rising investor activity, and possibly a shift towards more speculative trading behavior in the later years.

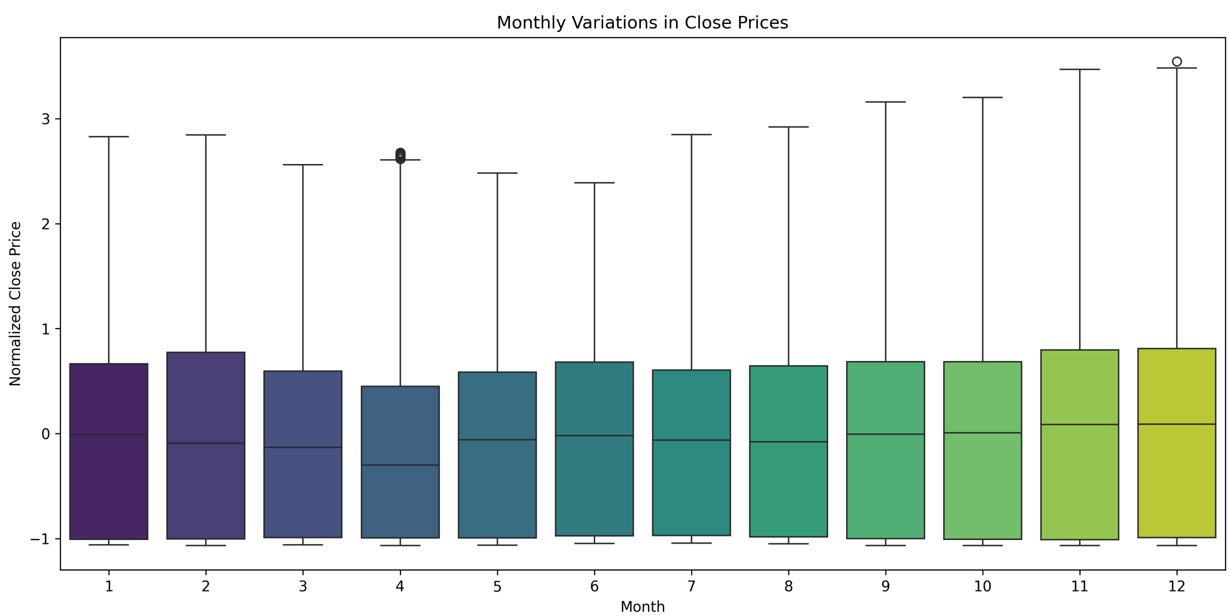
- **Trading Volume:** Trading Volume shows extreme spikes during the early years (1996–2000), followed by a significant decline and stabilization post-2004. A few sharp spikes are also observed around 2008 and 2020. *Significance:* High volumes in the late 1990s could reflect market bubbles or speculative trading activities. The spikes around 2008 and 2020 likely correspond to market events like the global financial crisis and pandemic, where increased trading activity indicates panic selling or opportunistic buying.

- **Volume-Price Relationship:** There appears to be no direct linear relationship between spikes in volume and changes in Close Prices. Large trading volumes do not always coincide with price surges or crashes. *Significance:* This suggests that while trading volume indicates market participation levels, price movements depend on broader factors such as sentiment, news, and fundamentals.

Overall Interpretation: The time series analysis of Close Prices and Trading Volume highlights a dynamic market with periods of growth, volatility, and high activity. Identifying these trends is critical for understanding market behavior, developing forecasting models, and designing effective hedging strategies.

3.4 Weekly and Monthly Variations in Close Prices





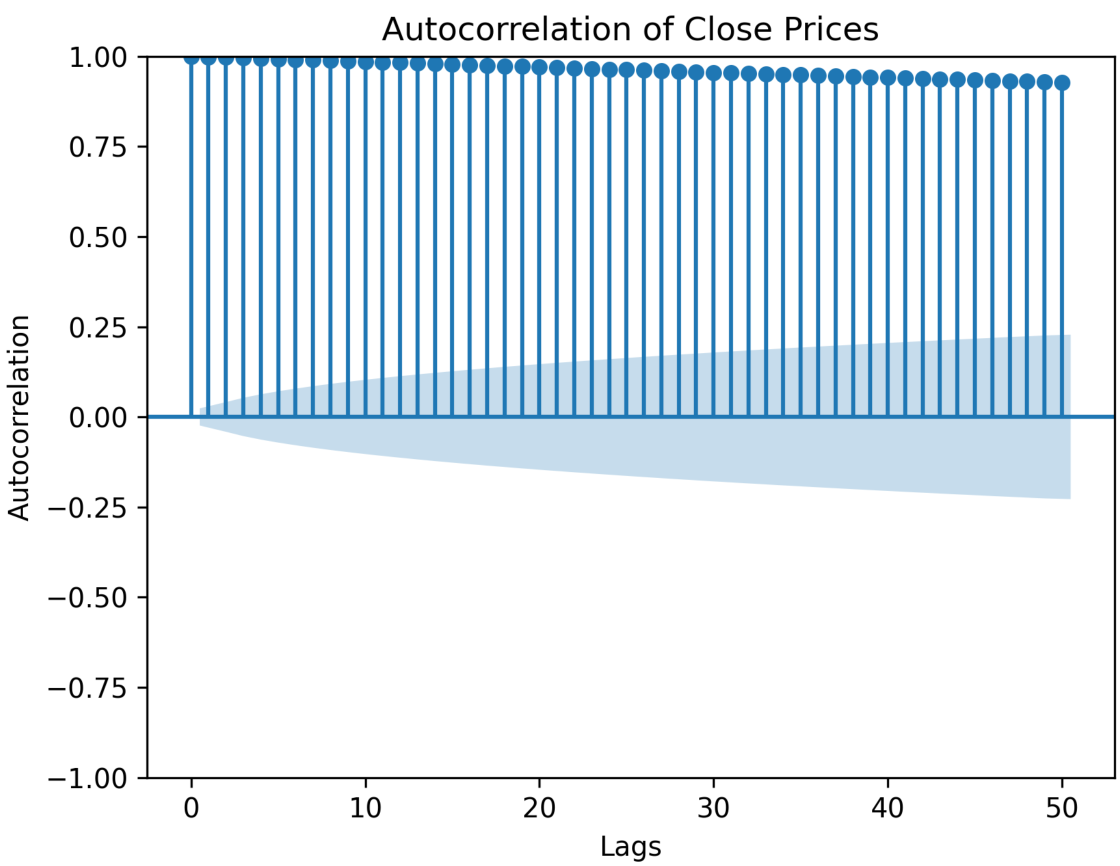
Observations and Significance:

- **Weekly Variations (Top Panel):** The median Close Prices remain relatively stable across the weeks, but the spread (indicated by box height and whiskers) varies significantly. Weeks towards the end of the year (weeks 40–52) display larger spreads and outliers. *Significance:* This may indicate increased market activity and volatility during the end of the year, often attributed to year-end portfolio adjustments, seasonal trading patterns, and tax-related activities.

- **Outliers in Weekly Data:** Weeks 13–15 and weeks 46–48 show a higher frequency of extreme values (outliers). *Significance:* These weeks might correspond to specific events, such as earnings announcements, policy changes, or external macroeconomic news triggering sharp price movements.
- **Monthly Variations (Bottom Panel):** The monthly variations in Close Prices show fairly consistent median values across all months. However, the spread increases notably in December (Month 12), with more extreme upper whiskers and outliers. *Significance:* December typically sees heightened volatility due to year-end trading strategies, including window dressing and tax-loss harvesting, where traders adjust their portfolios.
- **Volatility in April and December:** Month 4 (April) and Month 12 (December) display slightly larger whiskers, highlighting more volatile price movements. *Significance:* April often coincides with the end of financial years in certain economies, while December aligns with global year-end financial and trading activities.

Overall Interpretation: The analysis of weekly and monthly variations in Close Prices reveals that volatility tends to increase toward the end of the year, particularly in December. Identifying these seasonal trends helps in understanding market behavior, allowing traders and risk managers to prepare for periods of heightened volatility and optimize their hedging strategies accordingly.

3.5 Autocorrelation of Close Prices



Observations and Significance:

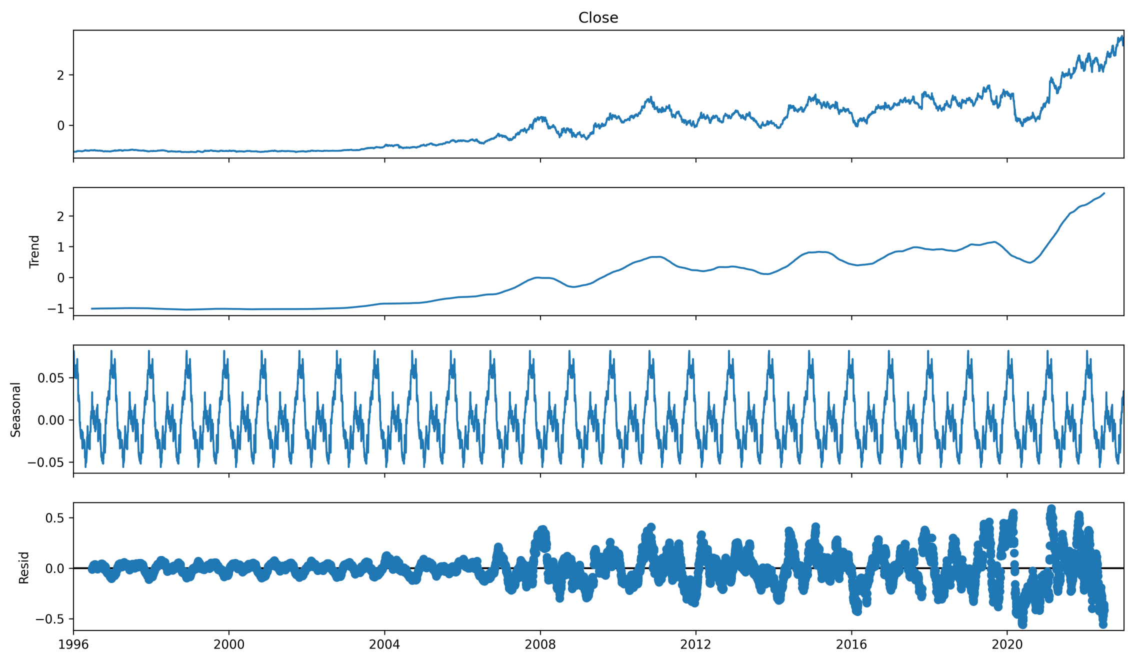
- **High Autocorrelation Across All Lags:** The ACF plot shows very high positive autocorrelation (close to 1) across all lags, with no significant decay over time. *Significance:* This indicates that Close

Prices exhibit strong persistence and follow a highly non-stationary process. The price at one time point is strongly dependent on the previous values.

- **Non-Stationarity of the Series:** The absence of decay in autocorrelation implies the series is not stationary. *Significance:* A non-stationary time series violates assumptions for many forecasting models, such as ARIMA. To make the data suitable for modeling, transformations such as differencing are required.
- **Long-Term Memory:** The slow decay of autocorrelation suggests the presence of long-term memory in the price series. *Significance:* This behavior is typical in financial time series, where trends persist over extended periods, making it crucial to account for such patterns in predictive models.

Overall Interpretation: The autocorrelation analysis reveals that the Close Price series is strongly correlated with its past values, suggesting non-stationarity and long-term trends. Appropriate preprocessing steps, such as differencing or detrending, will be required before applying time-series forecasting models to capture the underlying patterns effectively.

3.6 Trend, Seasonality, and Residual Decomposition of Close Prices



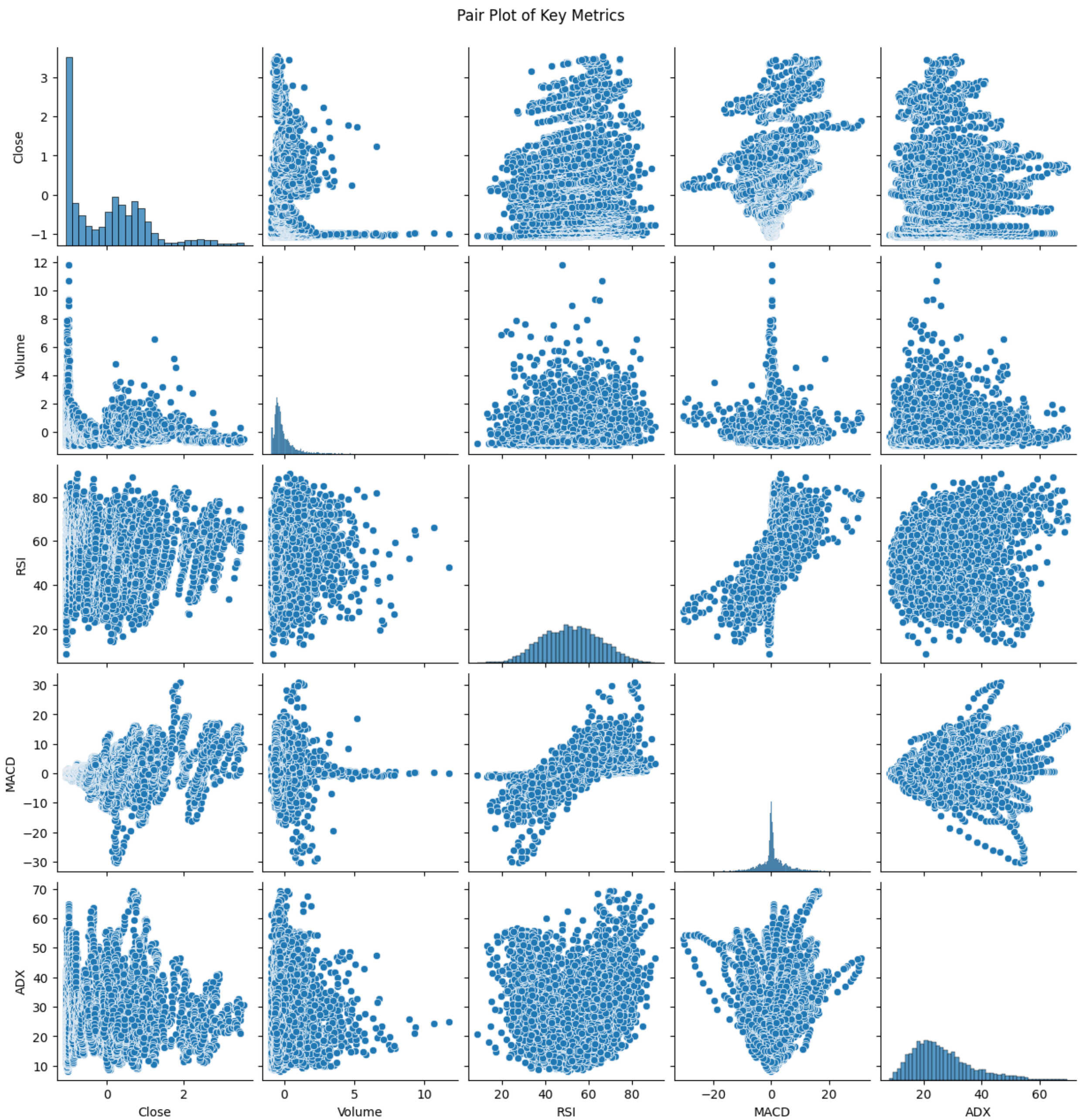
Observations and Significance:

- **Observed (Top Panel):** The original Close Prices time series shows a strong upward trend, with clear fluctuations and volatility after 2008. *Significance:* This upward trend indicates long-term growth in prices, though short-term volatility highlights the market's dynamic nature.
- **Trend (Second Panel):** The trend component highlights the smoothed, long-term growth of Close Prices. After a relatively flat period (1996–2004), a notable upward trend emerges, accelerating significantly post-2020. *Significance:* The increasing trend reflects fundamental market growth, possibly driven by macroeconomic factors, investor sentiment, and financial market conditions.
- **Seasonal (Third Panel):** The seasonal component shows repeating cyclical patterns over time. The amplitude of seasonal variations remains consistent, suggesting a stable seasonal influence on Close Prices. *Significance:* Such seasonality could be linked to periodic events, such as financial quarters, year-end trading, or other regular market cycles. Understanding these patterns is crucial for identifying profitable trading opportunities.

- **Residual (Bottom Panel):** The residual component captures the irregular fluctuations not explained by the trend or seasonality. Residuals show higher volatility post-2008, with increased variability after 2020. *Significance:* The higher residual volatility indicates the influence of unexpected shocks, such as market crashes, global events, or changes in investor behavior.

Overall Interpretation: The decomposition of Close Prices reveals a strong upward trend, consistent seasonality, and significant residual noise. This analysis helps separate long-term patterns from short-term fluctuations, providing insights for time series modeling and risk management. Accounting for these components improves the predictive power of forecasting models, especially in volatile markets.

3.7 Pair Plot of Key Metrics

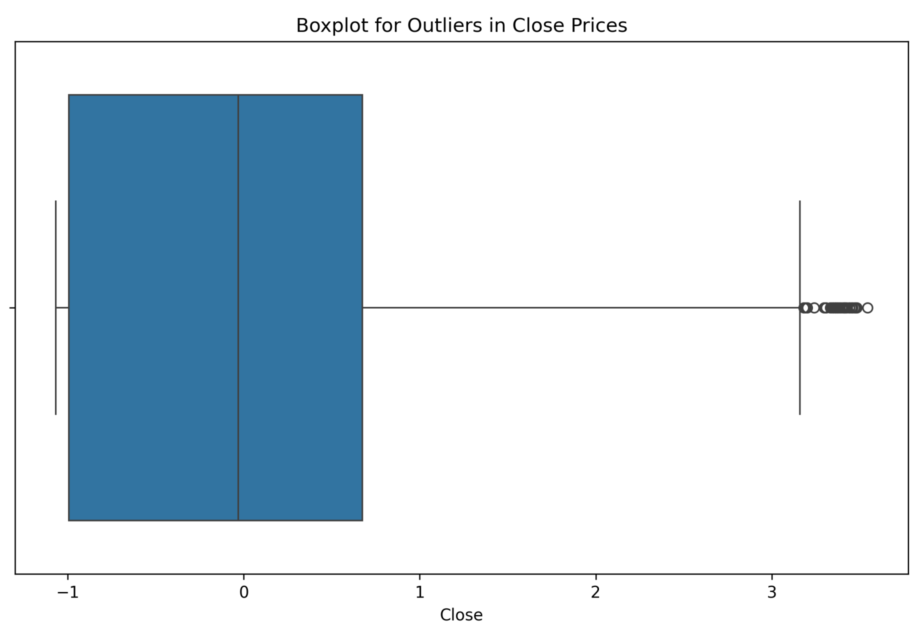


Observations and Significance:

- **Close Prices:** The univariate distribution of Close Prices is heavily skewed to the left, indicating the presence of smaller values with fewer higher values. *Significance:* Such skewness suggests price distributions typical in volatile or speculative markets.
- **Volume vs. Close Prices:** The scatter plot shows no clear linear relationship between Volume and Close Prices, though certain regions indicate clustering. *Significance:* Trading volume does not consistently predict price changes but highlights periods of higher market activity.
- **RSI vs. MACD:** A strong clustering pattern is visible in the RSI and MACD scatter plot, suggesting that these indicators are somewhat related, particularly in identifying overbought and oversold conditions. *Significance:* Both indicators can be complementary tools for traders to identify potential trend reversals.
- **ADX vs. Close Prices:** ADX (Average Directional Index) shows a spread pattern with Close Prices but no strong correlation. Higher ADX values suggest stronger trends regardless of direction. *Significance:* ADX helps quantify the strength of a trend, not its direction, which aligns with its weak relationship to Close Prices.
- **MACD vs. Close Prices:** MACD values are spread across a range of Close Prices, forming a distinct clustering pattern, especially at higher price levels. *Significance:* Since MACD is derived from moving averages, its patterns tend to align with significant price movements or trends.
- **Univariate Distributions:** - Close Prices: Skewed to the left. - Volume: Right-skewed with a few extreme peaks. - RSI: A relatively uniform distribution with slight clustering near overbought (70) and oversold (30) values. - MACD: Centered around zero, indicating frequent crossovers. - ADX: Concentrated below 40, signifying generally weaker trend strengths.

Overall Interpretation: The pair plot reveals key relationships between trading indicators and prices. While some features, like MACD and RSI, show clustering relationships, others, like Volume and Close Prices, appear weakly correlated. Understanding these pairwise relationships is essential for feature selection and constructing predictive models.

3.8 Boxplot for Outliers in Close Prices



Observations and Significance:

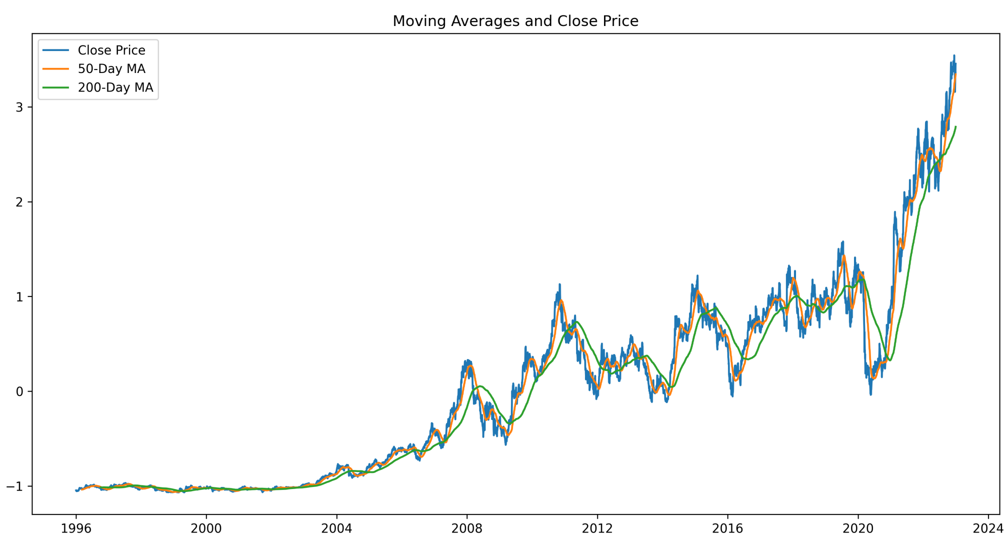
- **Presence of Outliers:** The boxplot reveals a significant number of outliers on the upper side of the Close Prices distribution. These outliers lie beyond the whiskers, which represent 1.5 times the interquartile range (IQR). *Significance:* Outliers often represent extreme price movements caused by sudden market events, such as financial news, economic shifts, or high speculative activity.

- **Skewed Distribution:** The median value of Close Prices is closer to the lower quartile, indicating a left-skewed distribution with more lower values. *Significance:* The skewness suggests that while most prices are relatively low, a small number of extreme values pull the distribution's tail upwards.

- **Impact on Analysis:** The presence of outliers can significantly affect statistical measures such as mean and variance. *Significance:* For time series modeling or risk management, these outliers need to be carefully handled. Methods like Winsorization, transformation, or robust statistical models can be applied to mitigate their impact.

Overall Interpretation: The boxplot highlights the non-symmetric nature of Close Prices and the presence of extreme upper outliers. These price anomalies may signal periods of market stress, volatility, or speculative activity, and require further investigation to determine their underlying causes.

3.9 Moving Averages and Close Price



Observations and Significance:

- **Close Price vs. Moving Averages:** The Close Price fluctuates around the 50-Day and 200-Day Moving Averages, with the 50-Day MA (orange line) being more responsive to short-term price movements, while the 200-Day MA (green line) reflects long-term trends. *Significance:* Moving averages smooth out noise in price data, helping to identify trends. Short-term MAs (e.g., 50-Day) are useful for detecting recent momentum, whereas long-term MAs (e.g., 200-Day) reveal broader market direction.

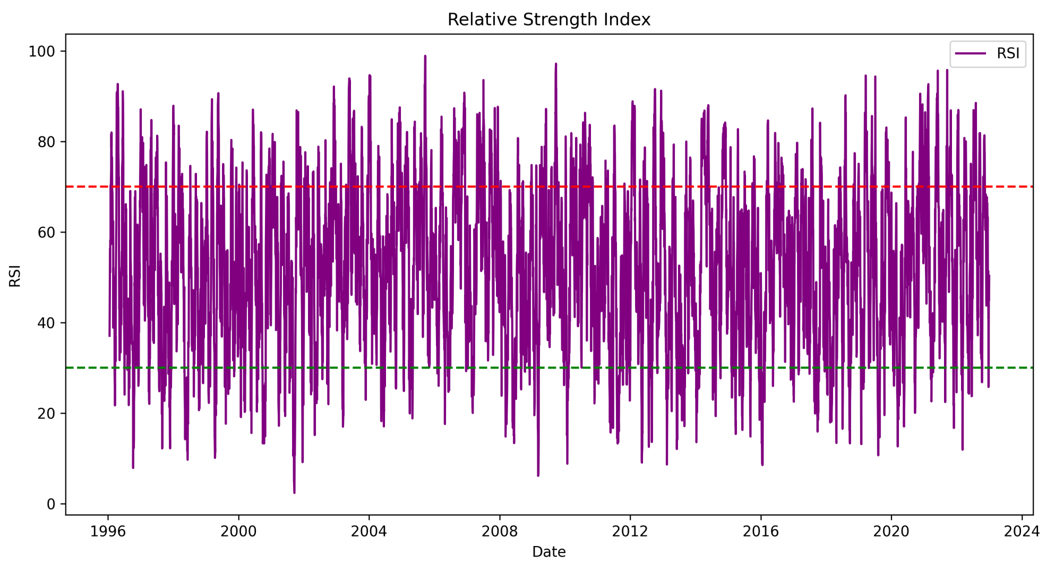
- **Golden Cross and Death Cross:** - A **Golden Cross** occurs when the 50-Day MA crosses above the 200-Day MA, signaling a potential upward trend (e.g., post-2020). - A **Death Cross** occurs when the 50-Day MA crosses below the 200-Day MA, indicating a bearish trend (e.g., during 2008). *Significance:* These crossover points are widely used as signals for trend reversals and are critical for timing entry and exit positions in trading strategies.

- **Market Cycles:** The plot shows distinct cycles of price increases and declines, especially around significant periods such as the 2008 financial crisis and the 2020 pandemic. *Significance:* Identifying these cycles helps investors and traders anticipate market movements and adjust hedging strategies to mitigate risks.

- **Post-2020 Growth:** The sharp rise in Close Prices post-2020 aligns closely with the upward movement of both moving averages. The 50-Day MA remains consistently above the 200-Day MA during this period, indicating a sustained bullish trend. *Significance:* Such alignment between moving averages confirms strong positive momentum, helping investors identify periods of market recovery.

Overall Interpretation: The combination of Close Prices and moving averages highlights the utility of trend-following indicators in analyzing market behavior. The 50-Day and 200-Day MAs are particularly useful for identifying short-term momentum and long-term market direction, respectively. Crossovers between these MAs act as key signals for potential trend reversals, which are critical for designing effective trading and hedging strategies.

3.10 RSI Chart



Observations and Significance:

- **Overbought and Oversold Levels:** The red dashed line at 70 marks the overbought threshold, and the green dashed line at 30 represents the oversold threshold. RSI values frequently oscillate between these levels. *Significance:* RSI values above 70 suggest the asset might be overbought, potentially signaling a price reversal or correction. RSI values below 30 indicate the asset might be oversold, suggesting a possible buying opportunity.

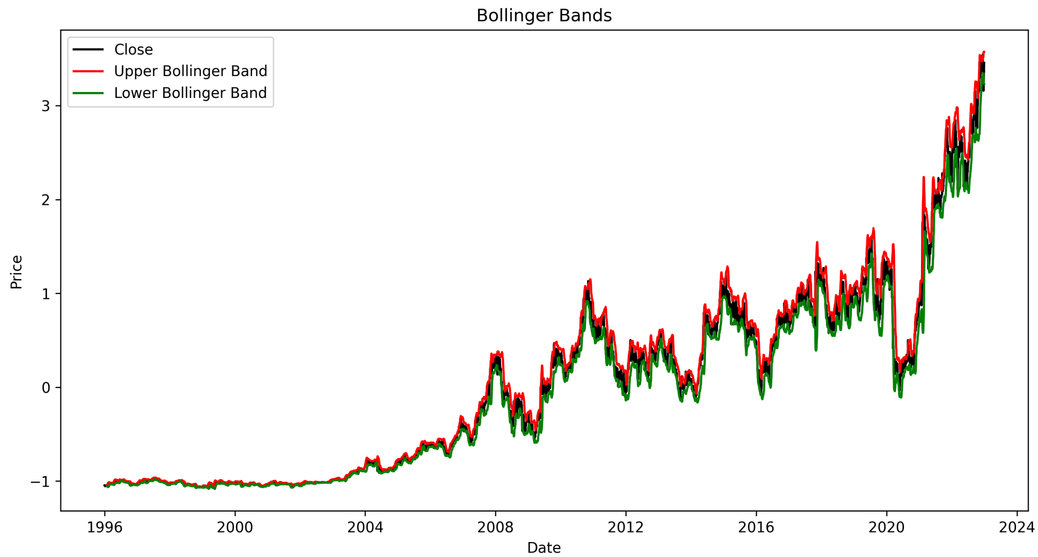
- **RSI Volatility:** The RSI exhibits significant volatility, fluctuating sharply over time. *Significance:* This volatility reflects the dynamic nature of market movements, as RSI is highly sensitive to recent price changes.

- **Sustained Overbought or Oversold Conditions:** RSI occasionally remains near or above 70 for extended periods, particularly during strong uptrends. Similarly, prolonged periods below 30 are rare, indicating a relatively balanced market over the long term. *Significance:* Sustained RSI extremes confirm strong momentum in the prevailing trend, which can help traders avoid premature reversals.

- **Market Cycles and Trend Reversals:** Sharp RSI drops below 30 or rises above 70 often coincide with turning points in price trends. Identifying these signals can aid in market entry or exit decisions. *Significance:* RSI is widely used in technical analysis to identify trend exhaustion and potential reversals, making it a critical indicator for short-term trading strategies.

Overall Interpretation: The Relative Strength Index (RSI) effectively highlights overbought and oversold market conditions. The frequent oscillation between the 30 and 70 thresholds reflects dynamic price momentum, while extreme values serve as potential signals for trend reversals. RSI remains a valuable tool for identifying entry and exit points in trading strategies.

3.11 Bollinger Bands Chart



Observations and Significance:

- **Bollinger Bands and Volatility:** The upper (red) and lower (green) Bollinger Bands expand and contract over time based on price volatility. Periods of higher volatility correspond to wider bands, while calmer periods result in narrower bands. *Significance:* The dynamic width of the Bollinger Bands highlights market volatility, enabling traders to identify periods of market stress or calm.

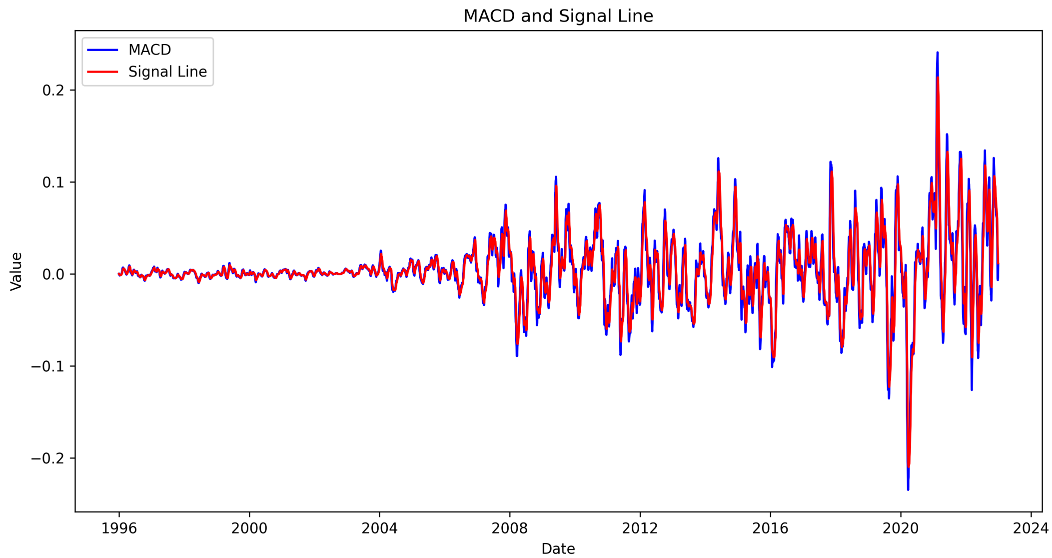
- **Price Interaction with Bands:** - Prices frequently approach or cross the upper Bollinger Band during strong uptrends, indicating potential overbought conditions. - Similarly, prices nearing or breaching the lower Bollinger Band often signal oversold conditions or a potential reversal. *Significance:* These price-band interactions are critical for traders to detect opportunities for trend reversals or continuations.

- **Periods of Market Trends:** During sustained uptrends (e.g., post-2020), prices consistently stay near the upper Bollinger Band, reflecting strong positive momentum. Conversely, during market corrections (e.g., 2008), prices move closer to or below the lower band. *Significance:* Bollinger Bands provide insights into trend strength and can confirm directional moves or reversals.

- **Squeeze Effect:** Bollinger Bands narrow significantly in certain periods (e.g., before 2004), which indicates low volatility. These squeezes are often followed by significant price movements (breakouts) in either direction. *Significance:* The squeeze effect helps identify potential breakout opportunities, making it a useful tool for traders and risk managers.

Overall Interpretation: Bollinger Bands are an effective technical analysis tool for understanding price volatility and potential trend reversals. Expansions and contractions in the bands reflect changing market dynamics, while price interactions with the bands provide actionable signals for traders to optimize their entry and exit points.

3.12 MACD Chart



Observations and Significance:

- **MACD and Signal Line Behavior:** The MACD (blue line) fluctuates around the zero line, while the Signal Line (red line) closely follows the MACD with a slight lag. *Significance:* - Positive MACD values (above zero) indicate upward momentum in the price trend. - Negative MACD values (below zero) suggest downward momentum.

- **Crossovers Between MACD and Signal Line:** - When the MACD crosses above the Signal Line, it generates a **bullish signal**, indicating a potential upward trend. - Conversely, when the MACD crosses below the Signal Line, it generates a **bearish signal**, indicating a potential downward trend. *Significance:* These crossovers are widely used in technical analysis to identify entry and exit points for trading.

- **Increased Volatility Post-2008:** The amplitude of both the MACD and Signal Line increases significantly post-2008, reflecting higher market volatility. *Significance:* This trend aligns with periods of economic uncertainty and greater price fluctuations, emphasizing the importance of MACD as a volatility-sensitive indicator.

- **Zero Line Crossings:** - MACD crossing the zero line from below to above suggests strengthening bullish momentum. - Crossing from above to below indicates weakening bullish momentum or strengthening bearish momentum. *Significance:* Zero line crossings confirm larger trend changes and can complement crossover signals for stronger trade confirmations.

Overall Interpretation: The MACD and Signal Line are powerful tools for identifying momentum and trend reversals in asset prices. Crossovers and zero-line crossings provide actionable trading signals, while increased amplitude reflects higher market volatility. This analysis is crucial for developing robust trading and hedging strategies.

4 Derived Trading Strategies Based on EDA

The Exploratory Data Analysis (EDA) provides critical insights into the behavior of Close Prices, technical indicators, and their relationships. This section outlines key trading strategies derived from the observations and analyses performed.

4.1 Strategy 1: Bollinger Bands for Trend Reversal Identification

Concept: Bollinger Bands help identify overbought and oversold market conditions based on price movements relative to the bands.

- **Buy Signal:** When the Close Price breaches the lower Bollinger Band, indicating an oversold condition, and subsequently rebounds.
- **Sell Signal:** When the Close Price breaches the upper Bollinger Band, indicating an overbought condition, and shows signs of reversal.
- **Stop-Loss:** Use a stop-loss below the lower band (for buy trades) or above the upper band (for sell trades) to minimize risks.

Significance: This strategy leverages Bollinger Bands to identify extreme price conditions and potential reversals, making it suitable for mean-reversion trades.

4.2 Strategy 2: RSI for Overbought and Oversold Signals

Concept: The Relative Strength Index (RSI) is used to determine overbought and oversold conditions in the market.

- **Buy Signal:** When RSI drops below 30 (oversold condition) and then rises back above 30.
- **Sell Signal:** When RSI rises above 70 (overbought condition) and subsequently drops below 70.
- **Stop-Loss:** Place a stop-loss at recent lows for buy signals and recent highs for sell signals.

Significance: RSI provides reliable signals for traders to identify trend exhaustion points and potential reversals, especially in ranging markets.

4.3 Strategy 3: MACD Crossover for Momentum Trading

Concept: The Moving Average Convergence Divergence (MACD) and Signal Line crossovers help identify shifts in market momentum.

- **Buy Signal:** When the MACD line crosses above the Signal Line, signaling increasing bullish momentum.
- **Sell Signal:** When the MACD line crosses below the Signal Line, signaling increasing bearish momentum.
- **Confirmation:** Combine MACD signals with Bollinger Band breakouts or RSI conditions to improve accuracy.

Significance: MACD is a widely used indicator for momentum-based strategies and works effectively in trending markets.

4.4 Strategy 4: Moving Average Crossovers for Trend Identification

Concept: The 50-Day and 200-Day Moving Averages (MA) help identify long-term and short-term trends.

- **Golden Cross (Buy Signal):** When the 50-Day MA crosses above the 200-Day MA, indicating a bullish trend.
- **Death Cross (Sell Signal):** When the 50-Day MA crosses below the 200-Day MA, signaling a bearish trend.
- **Trade Management:** Hold positions until the opposite crossover occurs or combine with other indicators for confirmation.

Significance: Moving average crossovers are reliable for identifying trend direction and are widely used in medium to long-term trading strategies.

4.5 Strategy 5: Volume Analysis for Breakout Confirmation

Concept: Trading volume provides insights into the strength of price movements and potential breakouts.

- **Buy Signal:** A significant increase in volume accompanying a price breakout above resistance levels.
- **Sell Signal:** A large volume spike with a breakdown below key support levels.
- **Confirmation:** Use Bollinger Bands or MACD to validate breakout strength.

Significance: Volume analysis helps confirm the validity of breakouts and prevents false signals in low-volume conditions.

4.6 Strategy 6: Combining Indicators for Robust Signals

Concept: Combining multiple indicators enhances the reliability of trading signals and reduces false positives.

- **Example Strategy: - Buy Signal:** RSI below 30, MACD crossover (bullish), and Close Price nearing the lower Bollinger Band. - **Sell Signal:** RSI above 70, MACD crossover (bearish), and Close Price nearing the upper Bollinger Band.

Significance: Combining indicators increases signal robustness, providing traders with greater confidence in decision-making.
Conclusion: The derived trading strategies leverage insights from Bollinger Bands, RSI, MACD, moving averages, and volume analysis. By combining these technical indicators, traders can develop robust, data-driven strategies to identify market trends, reversals, and breakout opportunities while managing risks effectively.

5 Feature Selection

Feature selection is a critical step in identifying the most relevant variables that contribute to the prediction of the target variable (*Close Price*). To ensure robustness and accuracy, three methods were applied: Correlation Analysis, Recursive Feature Elimination (RFE) with Random Forest, and Mutual Information (MI). The final set of features was derived by identifying common features selected by all three methods.

5.1 Feature Selection using Correlation Analysis

Method: Correlation analysis computes the linear relationship between each feature and the target variable (*Close Price*). Features with an absolute correlation coefficient greater than a predefined threshold (0.5) were selected for further analysis.
Results: The features with high positive or negative correlation with *Close Price* were identified as significant. These features were retained for further modeling.

5.2 Feature Selection using Recursive Feature Elimination (RFE)

Method: Recursive Feature Elimination (RFE) ranks features by their importance using a Random Forest Regressor as the estimator. RFE iteratively removes the least important features until the top 10 most relevant features are selected.
Results: The RFE method selected the top 10 features that contribute the most to the prediction of *Close Price* based on feature importance derived from the Random Forest model.

5.3 Feature Selection using Mutual Information

Method: Mutual Information (MI) measures the dependency between each feature and the target variable. Features with the highest mutual information scores were ranked, and the top 10 features were selected.
Results: The Mutual Information method identified the 10 most informative features based on their relationship with *Close Price*.

5.4 Final Selected Features

To ensure consistency and robustness, the final set of features was derived by identifying the intersection of the features selected by all three methods (Correlation Analysis, RFE, and Mutual Information).
Common Selected Features: The following features were consistently selected across all three methods:

- `lag1_close`
- `Lower Bollinger Band`
- `rolling_mean_close`
- `High`
- `Open`
- `Upper Bollinger Band`
- `wk_close`
- `Low`

Conclusion: By combining the results of Correlation Analysis, Recursive Feature Elimination, and Mutual Information, we identified a robust set of features that are most relevant for predicting the *Close Price*. These features will be used as input for subsequent predictive modeling and strategy development.

6 Model Development: LSTM for Time-Series Forecasting

In this section, we describe the development of an LSTM (Long Short-Term Memory) model to forecast the *Close Price* using the selected features. Hyperparameter tuning was performed using Keras Tuner to optimize the model performance.

6.1 Data Preparation and Feature Selection

Input Features: The following features, identified during the feature selection process, were used as inputs:

- `lag1_close`
- `Lower Bollinger Band`
- `rolling_mean_close`
- `High`

- `Open`
- `Upper Bollinger Band`
- `wk_close`
- `Low`

The *Close Price* was used as the target variable. Missing values were dropped, and all features were scaled using the Min-Max Scaler to normalize data between 0 and 1.

6.2 Sequence Preparation and Train-Test Split

To account for the temporal dependencies in time-series data, sequences of length 50 were created, where each sequence predicts the *Close Price* at the next time step. The data was split into 80% for training and 20% for testing.

6.3 LSTM Model Architecture and Hyperparameter Tuning

The LSTM model was implemented with the following architecture:

- **LSTM Layer:** A single LSTM layer with tunable units ranging from 32 to 128.
- **Dropout Layer:** A dropout layer with a tunable rate between 0.2 and 0.5 to prevent overfitting.
- **Dense Layer:** A fully connected Dense layer with a single output node to predict the target value.

Hyperparameter tuning was performed using the Keras Tuner library with the following configurations:

- **Objective:** Minimize validation loss (`val_loss`).
- **Tuning Space:**
  - **LSTM Units:** {32, 64, 96, 128}
  - **Dropout Rate:** [0.2, 0.3, 0.4, 0.5]
  - **Learning Rate:** {0.001, 0.0001}
- **Max Trials:** 10 trials with 2 executions per trial.
- **Callback:** Early stopping with patience of 5 epochs to restore the best weights.

6.4 Model Performance and Results

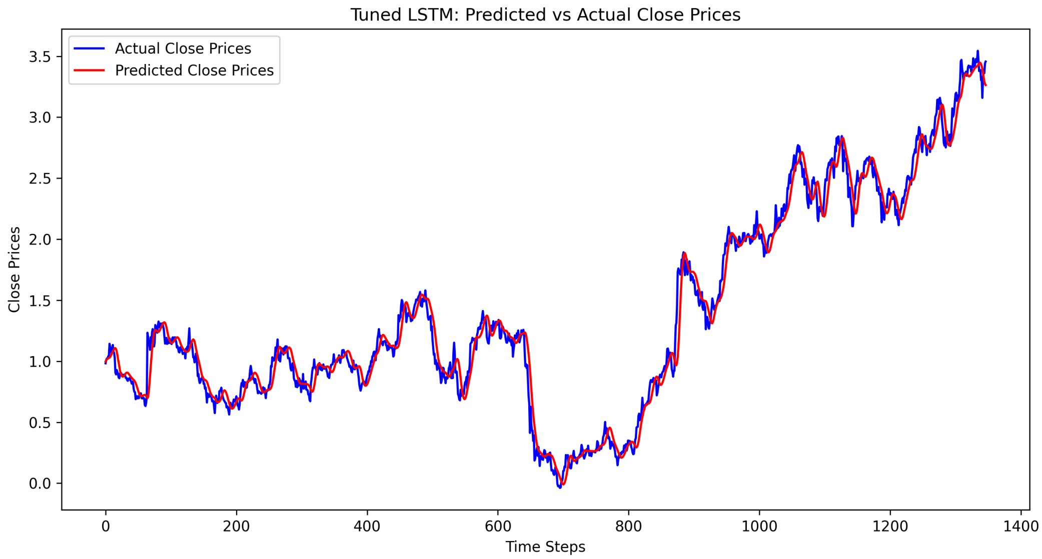
The hyperparameter search was completed in approximately 55 minutes, with the following best result:

- **Best Validation Loss:** 0.00076

The final tuned LSTM model was evaluated on the test dataset, and the following metrics were obtained:

- **Root Mean Squared Error (RMSE):** 0.1097
- **Mean Absolute Error (MAE):** 0.0804
- **Mean Absolute Percentage Error (MAPE):** 11.66%

Actual vs. Predicted Plot: The plot shows a comparison between the actual and predicted *Close Prices* on the test dataset.



Interpretation: The close alignment between actual and predicted values indicates that the LSTM model captures the trends and patterns in the data effectively.

6.5 Conclusion

The LSTM model, optimized through hyperparameter tuning, demonstrates strong predictive performance for forecasting the *Close Price*. The combination of selected features, sequence preparation, and hyperparameter optimization significantly contributed to minimizing prediction errors. The Actual vs. Predicted plot further validates the model's ability to generalize well on unseen data.

7 Trading Signal Generation

In this section, we describe the process of generating trading signals (BUY, SELL, HOLD) based on the actual and predicted prices obtained from the LSTM model. The signals provide actionable decisions for executing trades based on price movements.

7.1 Methodology

Signal Generation Logic: Trading signals were generated by analyzing the percentage change between the predicted price at time t and the actual price at time $t-1$. The decision thresholds were defined as follows:

- **BUY Signal:** If the percentage change exceeds a positive threshold of 1% (`THRESHOLD_UP` = 0.01).
- **SELL Signal:** If the percentage change is below a negative threshold of -1% (`THRESHOLD_DOWN` = -0.01).
- **HOLD Signal:** If the percentage change falls between -1% and +1%.

Mathematically, the signal is computed as:

$$\text{Change} = \frac{\hat{P}_t - P_{t-1}}{P_{t-1}}$$

where $\hat{P}_t$ is the predicted price at time t and P_{t-1} is the actual price at time $t-1$. Based on the computed change:

$$\text{Signal} = \begin{cases} \text{BUY} & \text{if Change} > \texttt{THRESHOLD\_UP} \\ \text{SELL} & \text{if Change} < \texttt{THRESHOLD\_DOWN} \\ \text{HOLD} & \text{otherwise.} \end{cases}$$

7.2 Results

Actual Prices	Predicted Prices	Signal
1.0116	1.0088	BUY
1.0229	1.0162	HOLD
1.0207	1.0213	HOLD
1.0393	1.0247	HOLD
1.0448	1.0271	SELL

Table 1: A Sample of Trading Signals Generated from LSTM Predictions

Interpretation: The generated signals provide clear trading decisions:

- A **BUY signal** indicates an anticipated upward movement, suggesting an opportunity to purchase.
- A **SELL signal** suggests an expected downward movement, recommending the sale of assets to mitigate losses or realize profits.
- A **HOLD signal** indicates minimal expected movement, suggesting no immediate action.

7.3 Conclusion

The trading signal generation process leverages the LSTM model's predictions and pre-defined thresholds to provide actionable BUY, SELL, or HOLD decisions. These signals form the basis for executing trades and designing automated trading strategies.

8 Conclusion

This report presented a comprehensive approach to developing a predictive hedging strategy using time-series forecasting techniques and trading signal generation. The primary objective was to predict *Close Prices* accurately and generate actionable trading signals to inform hedging decisions. Below are the key highlights of the study:

- **Exploratory Data Analysis (EDA):** EDA provided critical insights into data distribution, correlations between features, and patterns in price movements. Key technical indicators such as Bollinger Bands, RSI, and MACD were analyzed to understand trends, volatility, and potential reversals.
- **Feature Selection:** A robust feature selection process was conducted using three methods: Correlation Analysis, Recursive Feature Elimination (RFE), and Mutual Information (MI). The final selected features included:
  - `lag1_close`, `Lower Bollinger Band`, `rolling_mean_close`, `High`, `Open`, `Upper Bollinger Band`, `wk_close`, and `Low`.
- **LSTM Model Development:** An LSTM model was implemented for time-series forecasting, leveraging hyperparameter tuning to optimize performance. The model achieved the following metrics on the test dataset:
  - **Root Mean Squared Error (RMSE):** 0.1097

- **Mean Absolute Error (MAE):** 0.0804
- **Mean Absolute Percentage Error (MAPE):** 11.66%

The *Actual vs. Predicted* comparison demonstrated the model's effectiveness in capturing trends and patterns in price movements.

- **Trading Signal Generation:** A trading signal generation strategy was developed based on the LSTM model's predictions. Signals (BUY, SELL, HOLD) were generated using pre-defined thresholds, providing actionable decisions for trading and hedging.

Key Contributions: This study successfully demonstrated the integration of advanced time-series forecasting techniques with trading signal generation. The key contributions include:

- Accurate prediction of *Close Prices* using LSTM with optimized hyperparameters.
- Robust feature selection ensuring relevant and high-impact variables.
- Development of a systematic trading signal generation framework for hedging strategies.

Future Scope: Future enhancements can include:

- Incorporating additional external factors such as macroeconomic indicators and news sentiment analysis.
- Implementing ensemble models to further improve prediction accuracy.
- Deploying the developed model and trading signals on a real-time platform for automated hedging strategies.

Conclusion: The integration of LSTM-based time-series forecasting with trading signal generation provides a powerful and systematic approach to predicting price movements and informing hedging strategies. The results validate the model's predictive accuracy and its ability to generate actionable trading insights, demonstrating its applicability to real-world financial markets.

9 Future Work

While the current study successfully integrated time-series forecasting using LSTM with trading signal generation for hedging strategies, there are several areas for future improvements and extensions. The following directions are proposed for future work:

1. **Incorporation of External Factors:** Extend the current model by integrating external variables, such as:
   - **Macroeconomic Indicators:** Interest rates, inflation data, GDP growth, and unemployment rates.
   - **News Sentiment Analysis:** Real-time news feeds, sentiment scores from financial articles, and social media trends (e.g., Twitter data).
   - **Global Market Indicators:** Correlation with major indices, commodities, and forex markets.

   Incorporating these factors can help improve the robustness and adaptability of the forecasting model.

2. **Implementation of Ensemble Models:** Combine multiple machine learning models, such as LSTM, GRU (Gated Recurrent Units), and CNN (Convolutional Neural Networks), to form an ensemble framework. Ensemble models can leverage the strengths of different architectures to enhance prediction accuracy and reduce variance.

3. **Real-Time Deployment:** Deploy the developed model on a real-time platform for automated trading and hedging. Key steps include:

- Building a pipeline for continuous data ingestion and preprocessing.
- Implementing model retraining mechanisms to adapt to evolving market conditions.
- Integrating trading signals into a live trading system using APIs from brokers or exchanges.

4. **Optimization of Trading Strategies:** Further optimize the thresholds for trading signals (*BUY*, *SELL*, *HOLD*) using backtesting frameworks and reinforcement learning techniques. Additionally, explore position-sizing strategies to minimize risk and maximize returns.

5. **Risk Management and Portfolio Optimization:** Incorporate advanced risk management techniques, such as Value at Risk (VaR) and Conditional VaR, to quantify and mitigate risks. Combine trading signals with portfolio optimization algorithms (e.g., Markowitz's Mean-Variance Optimization) to construct efficient portfolios.

6. **Integration of Blockchain-Based Smart Contracts:** Leverage blockchain technology to automate hedging operations using smart contracts. The predictive signals can trigger real-time smart contracts to execute trades securely, transparently, and efficiently.

7. **Evaluation of Alternative Architectures:** Explore advanced deep learning models, such as:
   - **Transformers:** Use attention-based Transformer models, such as the Temporal Fusion Transformer (TFT), for time-series forecasting.
   - **Hybrid Models:** Combine statistical models (e.g., ARIMA or GARCH) with deep learning models for improved performance.

Conclusion: The proposed future work aims to enhance the accuracy, scalability, and practicality of the developed predictive hedging framework. By incorporating external factors, deploying real-time trading systems, and leveraging advanced models and technologies, the framework can be extended to meet the demands of dynamic and volatile financial markets.

References

- BDA. Big data analytics. Available at: *http://cs.rkmvu.ac.in/academics-msc-in-big-data-analytics-data-science/*, 2016.
- Fattah, Jamal; Ezzine, Latifa; Aman, Zineb; Moussami, Haj; Lachhab, Abdeslam. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, 10, 184797901880867. doi:10.1177/1847979018808673.
- Al Najjar, Dana. (2016). Modelling and Estimation of Volatility Using ARCH/GARCH Models in Jordan's Stock Market. *Asian Journal of Finance & Accounting*, 8, 152. doi:10.5296/ajfa.v8i1.9129.
- Olah, C. (2015). Understanding LSTMs. Available at: *https://colah.github.io/posts/2015-08-Understanding-LSTMs/*.