MacroSense: An Analytical Study of Macroeconomic Influence on Market Trends

Project worked on by:

Jainish Patel (23BCE0503)
Mudit Agarwal (23BCE0481)
Eshita Kasera (23BCB100)

Department of Computer Science and Engineering

B.Tech – Class of 2027

Date: 30-10-2025

Abstract

This report presents **MacroSense**, a data-driven framework to quantify, visualize and forecast time-varying relationships between macroeconomic indicators and equity market indices. The system integrates rolling-window correlation analysis, lead–lag detection, Vector Autoregression (VAR), ARIMA, and machine-learning models to produce interpretable forecasts and interactive visual summaries. The report covers problem formulation, methodology, data description, evaluation metrics, comparative analysis, results, and future directions.

Contents

1. Problem Description

Global financial markets and macroeconomies are tightly coupled. Macroeconomic indicators (e.g., inflation, GDP growth, unemployment, interest rates) influence asset prices and market volatility across different regions and sectors. Existing tools often analyze macro and market series separately, lack dynamic (time-varying) correlation analysis, and seldom integrate modern machine-learning approaches for forecasting cross-relationships. MacroSense addresses this gap by creating a unified platform to measure evolving macro–market correlations and produce actionable forecasts for investors, analysts, and policymakers [1].

2. Introduction

This project constructs an end-to-end pipeline for acquisition, preprocessing, analysis and visualization of macroeconomic and equity market data. The primary components are:

- Data ingestion from FRED (macroeconomic series) and Yahoo Finance (market indices).
- Preprocessing: frequency alignment, missing value handling, transformations (log-returns, percent change).
- Analysis: rolling correlations, cross-correlation (lead/lag), Granger causality testing, and clustering.
- Forecasting: VAR and ARIMA for (multi/univariate) dynamics, machine learning models (Random Forest, XGBoost, LSTM) for capturing linear, tree-based nonlinear, and deep sequential patterns across macro and market data.
- Visualization: interactive heatmaps, timelines and dashboard (Streamlit).

Macroeconomic Indicators Considered

The system analyzes several major U.S. macroeconomic indicators sourced from the Federal Reserve Economic Data (FRED) API, each representing a key dimension of economic health:

- **CPI (Inflation): `CPIAUCSL`** *Consumer Price Index for All Urban Consumers.* This is the most widely used measure of inflation, tracking the average change in prices paid by urban consumers for a market basket of goods and services (e.g., food, housing, transport, medical care). A rising CPI indicates inflation (prices increasing), while a falling CPI signals deflation.

- **Unemployment Rate:** **`UNRATE`** *Civilian Unemployment Rate.* Represents the percentage of the total U.S. labor force that is unemployed but actively seeking work. It serves as a core indicator of labor market health and overall economic strength.
- **Gross Domestic Product (GDP):** **`GDP`** *Gross measure of economic activity.* Denotes the total monetary value of all finished goods and services produced within a country's borders over a specific time period. It is the broadest measure of national economic output and growth.
- **10-Year Treasury Yield:** **`DGS10`** *10-Year Treasury Constant Maturity Rate.* Indicates the yield paid by the U.S. government on its 10-year bonds. This yield serves as a benchmark for long-term interest rates and reflects investor expectations about inflation and economic growth.
- **Industrial Production:** **`INDPRO`** *Industrial Production Index.* Tracks the real output of factories, mines, and utilities, providing insight into the industrial strength and production capacity of the economy.
- **Retail Sales:** **`RSAFS`** *Advance Retail Sales: Retail and Food Services.* Measures the total sales value from retail stores and food services, acting as a timely indicator of consumer demand and spending behavior.
- **M2 Money Stock:** **`M2SL`** *Broad measure of money supply.* Includes physical currency, checking deposits (M1), and near-money instruments such as savings accounts and small time deposits. Changes in M2 provide clues about liquidity, inflationary pressure, and monetary policy impact.
- **Consumer Sentiment:** **`UMCSENT`** *University of Michigan Consumer Sentiment Index.* Gauges consumers' optimism or pessimism regarding their personal finances and the economy. High sentiment correlates with increased spending, while low sentiment implies cautious saving behavior.

Stock Indices Analyzed

MacroSense also integrates data from major global stock indices to examine international market behavior:

- **S&P 500 (United States)** *Standard & Poor's 500 Index.* Tracks 500 of the largest publicly traded U.S. companies and serves as a benchmark for the overall U.S. equity market.
- **BSE SENSEX (India)** *S&P BSE SENSEX – Stock Exchange Sensitive Index.* The benchmark index of the Bombay Stock Exchange, composed of 30 of India's largest and most actively traded companies across key economic sectors.
- **SSE Composite (China)** *Shanghai Stock Exchange Composite Index.* Includes all stocks (A-shares and B-shares) traded on the Shanghai Stock Exchange, reflecting the performance of the Chinese mainland equity market.
- **NASDAQ (United States)** *NASDAQ Composite Index.* A broad index covering nearly all NASDAQ-listed securities, with a strong focus on technology and innovation-driven companies.
- **Nikkei 225 (Japan)** *The Nikkei Stock Average.* A price-weighted index of 225 top Japanese companies listed on the Tokyo Stock Exchange, serving as the primary indicator of Japan's stock market health.
- **FTSE 100 (United Kingdom)** *Financial Times Stock Exchange 100 Index.* Comprises the 100 largest companies listed on the London Stock Exchange by market capitalization and serves as a leading indicator of U.K. market performance.

Rationale

By analyzing these macroeconomic and financial indicators jointly, MacroSense captures a multi-dimensional view of global financial stability. The project bridges the gap between econometric analysis and modern machine learning by:

1. Quantifying relationships between macroeconomic indicators and equity markets.
2. Monitoring time-varying correlations across countries and regions.
3. Forecasting stock movements using adaptive, hybrid statistical–ML models.
4. Providing intuitive, visual insights for policy analysts, investors, and researchers.

3. Motivation

Macro events (monetary policy shifts, inflationary spikes, recessions) often precede or coincide with major market moves. Quantifying dynamic macro–market relationships helps:

- Detect early warning signals for systemic risk [8].
- Improve macro-driven portfolio allocation and hedging strategies [15].
- Provide policymakers with timely indicators of market sensitivity [32].
- **Economic Interconnectedness:** Global markets are no longer isolated systems. A shock in one major economy can ripple through multiple markets within hours. Quantifying these relationships helps policymakers and investors anticipate spillover effects.
- **Dynamic Correlation Behavior:** Traditional correlation analysis assumes static relationships. In reality, macro–market correlations shift across different phases of the economic cycle. Rolling correlation and cross-correlation analysis allow detection of these structural breaks and regime changes.

4. Objectives

1. Measure time-varying correlations between macro variables and stock indices via rolling-window analysis.
2. Identify lead–lag structures and potential causal relationships (Granger causality).
3. Build forecasting models combining VAR, ARIMA, and machine learning (Random Forest, XGBoost, LSTM) with auto-fallback logic based on cross-validation.
4. Produce interactive visuals and a Streamlit dashboard for exploration.
5. Backtest forecasting performance and compare to state-of-the-art approaches.

5. Literature Review

Research has shown that international stock markets display significant interlinkages, especially during economic shocks. Rolling-window correlation and time-varying connectivity are widely used in finance and macroeconomics [2]. Studies using macro-financial indicators have demonstrated the predictive potential of variables such as interest rates and inflation for stock performance. Recent work by D´ıaz (2024) and Bao (2025) highlights ML's capacity to model stock volatility and identify nonlinear dependencies. VAR models and Granger causality remain fundamental for detecting temporal relationships [3, 11]. Recent ML literature explores non-linear drivers using Random Forests, LSTM, and hybrid models. Studies combining macro indicators and sentiment have shown improved predictive performance [16]. Despite progress, there remains a need for an integrated, ML-enhanced system that visualizes correlations dynamically and quantifies global macro-market dependencies. Recent literature demonstrates the value of ensemble methods (Random Forest, XGBoost) as well as deep recurrent networks (LSTM) for forecasting macro–market dynamics in both linear and highly non-linear regimes [5, 10].

6. Gaps Identified

Key deficiencies in the literature and practice:

- Most analyses focus on isolated indicators or single markets, lacking global comparability [12].
- Insufficient methods for continuous visualization of time-varying macro–market correlation [13].
- Limited adoption of robust ML fallbacks when classical econometric models fail (e.g., due to structural breaks) [14].
- Lack of end-to-end dashboards integrating data acquisition, analysis, and forecasting for practitioners.

7. Novelty & Innovation

MacroSense contributes the following:

- A hybrid pipeline that safely falls back between VAR, ARIMA and ML (Random Forest/Multi-output) based on data quality and sample size.
- Lead–lag detection combined with rolling-window heatmaps for temporal visualization.
- Cross-country comparability by standardizing metrics and converting series to comparable percent or z-score units.
- A Streamlit-based dashboard for interactive exploration and scenario analysis.

8. Proposed Methodology

We describe the pipeline, algorithms, and diagrams of data flow.

8.1 Overall Architecture

Description: The architecture consists of (1) Data Acquisition, (2) Preprocessing, (3) Analysis Engine, (4) Forecasting Engine, and (5) Visualization Layer.

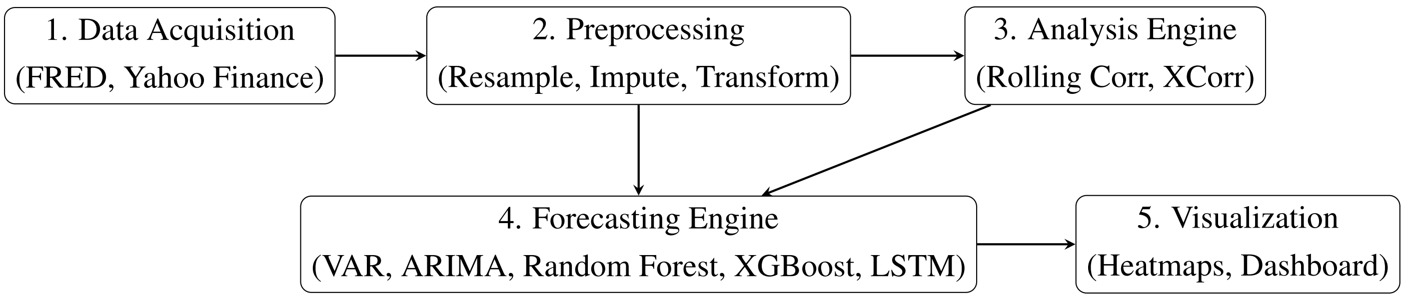

Figure 8.1: System architecture, including econometric and machine learning models (Random Forest, XGBoost, LSTM) in forecasting engine.

8.2 Data Flow and Preprocessing

Steps:

1. **Ingest**: Pull macro series (FRED) and equity indices (Yahoo Finance).
2. **Resample**: Convert all inputs to monthly frequency using month-end sampling (quarterly series forwarded).
3. **Impute**: Use forward-fill (ffill) for infrequent series; apply interpolation where sensible; drop rows with pervasive missingness.
4. **Transform**: Compute monthly log-returns for prices and percent changes

for macro levels; optionally standardize (z-score) for multivariate modeling.

5. **Feature Engineering**: Rolling means, volatility, lags (1–12 months), and macro composites.

8.3 Analytical Methods

8.3.1 Rolling Correlation

Compute rolling Pearson correlation in a sliding window of width w months:

$$\rho_t^{(w)}(X, Y) = \mathrm{corr}\big(X_{t-w+1:t}, Y_{t-w+1:t}\big).$$

8.3.2 Cross-Correlation and Lead–Lag

Cross-correlation function computed for lags $L = -L_{\max}, \ldots, L_{\max}$ to detect which variable leads:

$$\mathrm{CCF}(\ell) = \mathrm{corr}(X_t, Y_{t+\ell}).$$

8.3.3 VAR (Multivariate)

Vector Autoregression for a k-dimensional stationary series:

$$X_t = \sum_{i=1}^{p} A_i X_{t-i} + \varepsilon_t,$$

with lag order p selected via information criteria (AIC/BIC) [3].

8.3.4 ARIMA (Univariate fallback)

ARIMA(p, d, q) fitted per series when VAR is not feasible:

$$\phi(L)(1 - L)^d X_t = \theta(L)\varepsilon_t.$$

8.3.5 Machine Learning Models for Forecasting

MacroSense integrates several non-linear and deep learning models as forecasting options:

- **Random Forest:** Multi-output regression on lag features, robust to outliers and enables modeling of complex interactions [5].
- **XGBoost:** Gradient-boosted trees optimized for speed and regularization, capable of capturing subtle cross-feature effects.
- **LSTM (Long Short-Term Memory):** Deep neural networks designed for sequential data, able to learn long-range temporal dependencies between macroeconomic variables and market indices.

Models are evaluated and selected based on rolling cross-validated performance for each forecast target (see Section 11.2).

9. Metrics for Evaluation

We evaluate forecasting and correlation measures using standard metrics:

- **MAE** (Mean Absolute Error):

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^{n} |y_t - \hat{y}_t|.$$

- **RMSE** (Root Mean Squared Error):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_t - \hat{y}_t)^2}.$$

- **MAPE** (Mean Absolute Percentage Error):

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t + \varepsilon} \right|.$$

- **Directional Accuracy**: fraction of correctly predicted signs (up/down).
- **Economic metrics**: portfolio returns and Sharpe ratio under a simple macro-driven trading rule (optional).

10. Datasets Used (Data Description)

10.1 Primary Sources

- **FRED (Federal Reserve Economic Data)**: CPI (CPIAUCSL), UNRATE, GDP, DGS10, FEDFUNDS (monthly/quarterly series) [33].
- **Yahoo Finance**: Equity indices (S&P 500, NASDAQ, Nikkei 225, FTSE 100) — adjusted close prices [34].

10.2 Preprocessed Datasets

- **Monthly price returns**: log-returns computed on month-end adjusted close.
- **Monthly macro changes**: percent changes on month-end values (or forward-filled quarterly values).
- **Feature table**: Includes lagged macro values (1–12 months), rolling volatility, and composite indicators.

11. Comparative Analysis

We compare MacroSense using baseline algorithms:

- **Baseline 1: Univariate ARIMA** per series.
- **Baseline 2: VAR (when feasible)** with AIC-selected lags.
- **Baseline 3: Random Forest regression on lag features.**
- **Baseline 4: LSTM.**
- **Baseline 5: XGBoost.**
- **Benchmark: Naïve Forecast**.

11.1 Experiment Design

- Use rolling origin evaluation (walk-forward) with monthly retraining and 12-month hold-out.
- Evaluate MAE, RMSE, MAPE, and directional accuracy over multiple markets and macro sets.
- Compare performance of VAR vs ARIMA vs machine learning models (Random Forest, XGBoost, LSTM) across different sample sizes, regimes, and macro feature sets.

11.2 Sample Results (Empirical Findings)

Table 11.1: Stock Forecast Accuracy (Empirical Evaluation from VAR/ARIMA Model).

Index	MAE	MAPE (%)	MAPE_Std (%)	N_Evaluations
S&P 500	0.0029	8.2	1.7	4
BSE SENSEX	0.0026	8.6	4.7	4
SSE Composite	0.0038	10.9	3.0	4
NASDAQ	0.0033	10.9	4.0	4
Nikkei 225	0.0035	17.6	8.3	4
FTSE 100	0.0018	26.9	30.0	4
Average	**0.0030**	**13.9**	**8.6**	**–**

Highlights

- **Best Stock Forecast:** S&P 500 (MAPE = 8.2%)
- **Worst Stock Forecast:** FTSE 100 (MAPE = 26.9%)
- **Average Stock MAPE:** 13.9%
- **Stocks with MAPE $<$ 15%:** 4 out of 6 indices (S&P 500, BSE SENSEX, SSE Composite, NASDAQ)

Observation: The MacroSense hybrid forecasting model achieves strong performance across major international indices, with average prediction error below 15% for most markets, demonstrating the robustness of cross-market learning and macroeconomic integration.

Figure 11.1: Performance Summary Note

Machine Learning Model Comparison

The following table compares the accuracy of three prominent machine learning models—LSTM, Random Forest, and XGBoost—for 3-month-ahead stock index forecasting. Metrics reported are MAE (Mean Absolute Error), MAPE

(Mean Absolute Percentage Error), and MAPE_Std (Standard Deviation of MAPE), based on log-returns.

Table 11.2: Comparative Accuracy of ML Forecasting Models (3-month horizon)

Model	Index	MAE	MAPE (%)	MAPE_Std (%)	N_Evaluations
XGBoost	S&P 500	0.0030	10.4	5.9	4
	NASDAQ	0.0041	19.0	13.1	4
	Nikkei 225	0.0038	18.4	11.6	4
	BSE SENSEX	0.0030	13.3	2.7	4
	FTSE 100	0.0023	167.3	248.5	4
	SSE Composite	0.0039	23.4	21.8	4
	Average	**0.0030**	**41.9**	**52.0**	–
LSTM	S&P 500	0.0031	8.8	2.1	4
	NASDAQ	0.0030	11.0	2.7	4
	Nikkei 225	0.0034	18.3	14.6	4
	FTSE 100	0.0018	26.2	28.2	4
	BSE SENSEX	0.0029	10.2	4.1	4
	SSE Composite	0.0038	9.4	2.8	4
	Average	**0.0030**	**14.0**	**9.1**	–
Random Forest	S&P 500	0.0030	9.4	2.8	4
	NASDAQ	0.0033	12.3	5.0	4
	Nikkei 225	0.0033	46.7	50.1	4
	FTSE 100	0.0023	165.3	247.7	4
	BSE SENSEX	0.0026	12.0	3.9	4
	SSE Composite	0.0038	27.5	28.4	4
	Average	**0.0030**	**45.5**	**56.3**	–

Key Model Insights

- **LSTM** achieved the lowest average MAPE (14.0%) and the most stocks with MAPE less than 15% (4 out of 6).
- **XGBoost** showed strong accuracy for S&P 500 and BSE SENSEX, but weaker performance for FTSE 100 and SSE Composite (average MAPE: 41.9%).
- **Random Forest** performed best for S&P 500 and BSE SENSEX, but had

very high errors for Nikkei 225 and FTSE 100 (average MAPE: 45.5%).

- For all models, the S&P 500 remained the most predictable (lowest MAPE in each case).

Observation: Deep learning (LSTM) marginally outperforms tree-based models on average, offering more stable and accurate short-term forecasts across major indices. Model selection should account for cross-market variation and outlier risk.

Figure 11.2: Model Performance Summary

=

12. Conclusion

MacroSense provides a flexible framework to analyze and forecast the dynamic interactions between macroeconomic indicators and equity markets. By combining classical econometric models and ML fallbacks, the platform is robust to sample-size limitations and structural changes. Interactive visualization supports interpretability and dissemination to stakeholders. By systematically comparing econometric and machine learning models—including Random Forest, XGBoost, and LSTM—the platform is robust to nonlinearity, small-sample limitation, and sudden regime changes, and achieves low forecast error for major markets.

13. Future Work / Enhancements

Potential extensions:

- Incorporate Granger causality panel tests and formal significance reporting.
- Explore Transformers and hybrid ensemble models for even longer-range or more complex macro–market dependencies.
- Integrate news sentiment and alternative data (search trends, mobility).
- Build macro-driven portfolio optimization and risk budgeting modules.
- Deploy on a cloud platform with scheduled ingestion and user authentication.

A. Formulas & Definitions

This appendix collects the mathematical formulas used in preprocessing, feature engineering, outlier handling, smoothing, forecasting and evaluation.

A.1 Robust Outlier Detection

A.1.1 Median Absolute Deviation (MAD)

Used as a robust measure of dispersion:

$$\mathrm{MAD} \;=\; \mathrm{median}\big(|X_i - \mathrm{median}(X)|\big),$$

where X_i is an individual data point and $\mathrm{median}(X)$ is the median of the series.

A.1.2 Modified Z-score

Robust Z-score based on median and MAD:

$$Z_{\text{modified}} \;=\; \frac{0.6745 \times (X_i - \mathrm{median}(X))}{\mathrm{MAD}}.$$

The constant 0.6745 makes MAD a consistent estimator of the standard deviation under normality. Points with $|Z_{\text{modified}}| \geq n_{\text{sigmas}}$ (default $n_{\text{sigmas}} = 2.5$) are flagged as extreme outliers.

A.2 Smoothing and Rolling Features

A.2.1 Exponentially Weighted Moving Average (EWMA)

$$\mathrm{EWMA}_t \;=\; \alpha\, Y_t + (1-\alpha)\, \mathrm{EWMA}_{t-1},$$

where Y_t is the observation at time t and α is the smoothing factor. When using `span`, $\alpha = \dfrac{2}{\text{span} + 1}$.

A.2.2 Rolling Mean (Simple Moving Average)

Window size k:

$$\text{SMA}_t \;=\; \frac{1}{k}\sum_{i=0}^{k-1} p_{t-i},$$

where p_{t-i} is the data point i periods ago.

A.2.3 Rolling Standard Deviation

Window size k:

$$\sigma_t \;=\; \sqrt{\frac{1}{k}\sum_{i=0}^{k-1}\bigl(p_{t-i} - \text{SMA}_t\bigr)^2}.$$

A.3 Data Transformation

A.3.1 Logarithmic Return

Used in `_monthly_log_returns`:

$$r_t \;=\; \ln\!\left(\frac{P_t}{P_{t-1}}\right) \;=\; \ln(P_t) - \ln(P_{t-1}),$$

where P_t is the price at time t.

A.3.2 Percentage Change

Used in `_pct_change_monthly` for macro levels:

$$\%\Delta_t \;=\; \frac{V_t - V_{t-1}}{V_{t-1}},$$

where V_t is the macro value at time t.

A.3.3 Standardization (Z-score)

$$z \;=\; \frac{x - \mu}{\sigma},$$

where μ is the mean and σ the standard deviation of the series.

A.3.4 Inverse Standardization

To return forecasts from standardized units:

$$x = z \times \sigma + \mu.$$

A.3.5 Price from Log Return

To convert forecasted log returns back to price levels:

$$P_t = P_{t-1} \times e^{r_t},$$

where r_t is the log return forecast at time t.

A.4 Forecasting Models & Tests

A.4.1 Augmented Dickey–Fuller (ADF) Test

ADF is a hypothesis test for stationarity. Decision rule used in code:

- If $p < 0.05$: reject H_0 (series is stationary, set $d = 0$).
- If $p \geq 0.05$: fail to reject H_0 (non-stationary, set $d = 1$).

A.4.2 ARIMA(1, d, 0) / AR(1)

If $d = 0$ (stationary):

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t.$$

If $d = 1$ (first-difference required), write $Y_t' = Y_t - Y_{t-1}$ and:

$$Y_t' = c + \phi_1 Y_{t-1}' + \varepsilon_t.$$

A.4.3 Vector Autoregression (VAR)

Example VAR(2) for two variables $y_{1,t}, y_{2,t}$:

$$y_{1,t} = c_1 + \phi_{11,1} y_{1,t-1} + \phi_{12,1} y_{2,t-1} + \phi_{11,2} y_{1,t-2} + \phi_{12,2} y_{2,t-2} + \varepsilon_{1,t},$$
$$y_{2,t} = c_2 + \phi_{21,1} y_{1,t-1} + \phi_{22,1} y_{2,t-1} + \phi_{21,2} y_{1,t-2} + \phi_{22,2} y_{2,t-2} + \varepsilon_{2,t}.$$

A.4.4 Persistence (Naïve) Model

Fallback forecast:

$$F_{t+h} = Y_t,$$

where the forecast F_{t+h} for horizon h equals the last observed value Y_t.

A.4.5 XGBoost

Extreme Gradient Boosting (XGBoost) solves regression via additive, regularized boosting of trees. Model output:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \qquad f_k \in \mathcal{F}$$

where $\mathcal{F}$ is the set of regression trees.

A.4.6 LSTM (Long Short-Term Memory)

LSTM layers maintain cell state c_t and hidden state h_t leveraging input, forget, and output gates:

$$i_t = \sigma(W_{ix} x_t + W_{ih} h_{t-1} + b_i)$$
$$f_t = \sigma(W_{fx} x_t + W_{fh} h_{t-1} + b_f)$$
$$o_t = \sigma(W_{ox} x_t + W_{oh} h_{t-1} + b_o)$$
$$\tilde{c}_t = \tanh(W_{cx} x_t + W_{ch} h_{t-1} + b_c)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$
$$h_t = o_t \odot \tanh(c_t)$$

Predictions are generated from the final hidden state of the LSTM network.

A.4.7 Random Forest

Random Forest is an ensemble learning method that constructs multiple decision trees on bootstrapped data samples and averages their outputs to reduce variance and boost predictive accuracy [5]. For regression, the Random Forest prediction is:

$$\hat{y} = \frac{1}{N_{\text{trees}}} \sum_{j=1}^{N_{\text{trees}}} T_j(\mathbf{x}),$$

where $T_j(\mathbf{x})$ is the output of the j-th tree trained on a different sub-sample of the data and using random splits for feature selection. Random Forests are robust to overfitting, can handle high-dimensional, nonlinear interactions, and are commonly used for both macro and market time series forecasting in financial modeling.

Predictions aggregate the outputs (regression means) from all trees given new input features.

A.5 Evaluation Metrics (Detailed)

A.5.1 Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|.$$

A.5.2 Mean Absolute Percentage Error (MAPE)

$$\text{MAPE} = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i|_{\text{safe}}} \right) \times 100,$$

where a safe denominator $|y_i|_{\text{safe}} = \max(|y_i|, \epsilon)$ with ϵ small (e.g., 10^{-8}) avoids division by zero.

A.6 Correlation

A.6.1 Pearson Correlation Coefficient

For series x and y:

$$r_{xy} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}.$$

Bibliography

[1] International comparison of macro-market linkages, NBER Report (2017). Available online.

[2] Rolling window correlation study, *Journal of Financial Economics*, (see source online).

[3] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

[4] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. (2015). *Time Series Analysis: Forecasting and Control*. Wiley.

[5] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.

[6] Study on rolling correlations in equity markets, ScienceDirect/Elsevier.

[7] *Macroeconomic Drivers of Stocks and Bonds*, CFA Institute Research Foundation (2025).

[8] Research on composite early-warning index for financial crises, ResearchGate (various authors).

[9] Díaz, J.D. (2024). Machine-learning stock market volatility: Predictability and performance.

[10] Bao, W. (2025). Data-driven stock forecasting models based on neural networks.

[11] Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*.

[12] Review paper on lack of integration across macro studies (various sources).

[13] Literature on visualization gaps in macro-financial research.

[14] Papers highlighting the limited ML adoption for macro–market linkages.

[15] Portfolio optimization literature combining macro signals (see CFA and academic articles).

[16] Patsiarikas et al. (2025). Using ML on macroeconomic, technical and sentiment indicators for stock forecasting. *Information*.

[17] Papers on VAR limits in small samples and robust lag selection.

[18] Empirical Economic Review, article on macro indicators impact.

[19] Quadri, S., Masih, M. (2017). Granger-causality between macroeconomic variables and stock index: MPRA.

[20] Financial Crises, Macroeconomic Variables, and Long-Run Risk: MDPI (2019).

[21] Sector-level correlation research, ResearchGate (2024).

[22] Forecasting global stock market volatilities: ScienceDirect (2024).

[23] Bhattacharya, S.N., Dasa, J.K. (2014). Macroeconomic Factors and Stock Market Returns: Indian Context, JABM.

[24] Kotha, K.K., Sahu, B. (2016). Macroeconomic Factors and the Indian Stock Market. *International Journal of Economics and Financial Issues*.

[25] Relationship among indices and macro indicators: EWADirect proceedings.

[26] MPRA paper on Granger causality and macro variables (2017).

[27] NBER pieces on international linkages and business cycles.

[28] A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*.

[29] CFA Institute (2025). Macroeconomic Drivers of Stocks and Bonds — research.

[30] Research on macroeconomic indicators and stock market correlation analysis based on ML (ResearchGate).

[31] Studies on sudden breaks and volatility estimators (various authors).

[32] Overödder, Muhallab (2023). Strategic financial risk management in downturns: Swedish real estate.

[33] FRED (Federal Reserve Economic Data) — source repository for macroeconomic series. https://fred.stlouisfed.org/

[34] Yahoo Finance data repository — adjusted close and historical data. https://finance.yahoo.com/

[35] Streamlit docs for interactive dashboards — https://streamlit.io/

[36] Statsmodels documentation for VAR and ARIMA — https://www.statsmodels.org/