PROJECT PRESENTATION:
COFFEE PRICE FORCASTING

Course Project : **STATISTICS FOR COMPUTER SCIENCE CS309**

NOVEMBER 10TH 2025

ADARSH S KAMATAGI – 23BCS005

BABHALE ROHAN LAXMIKANT – 23BCS026

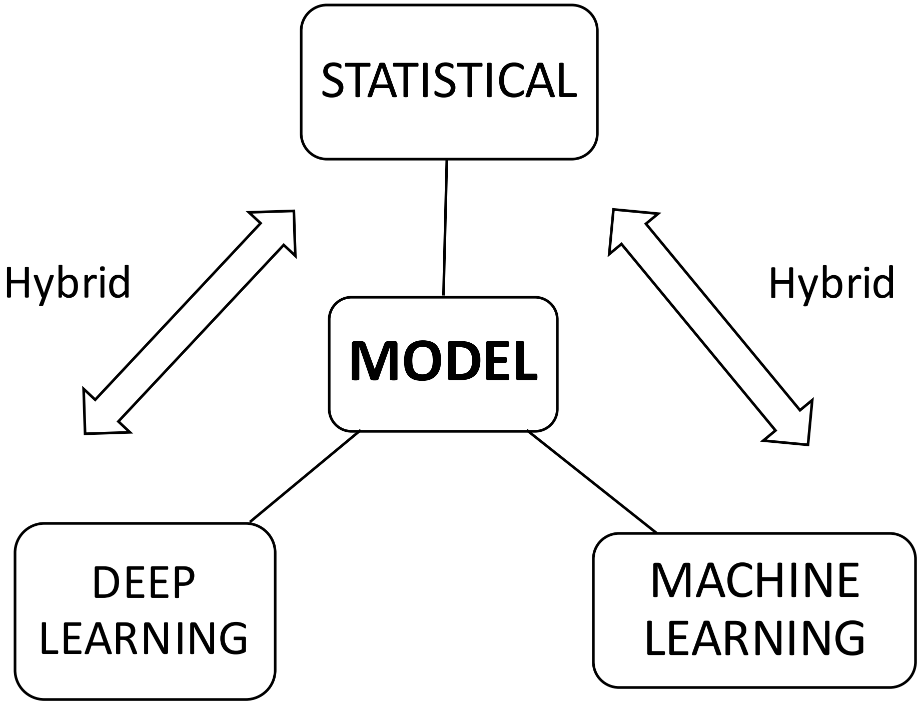
Why Coffee?

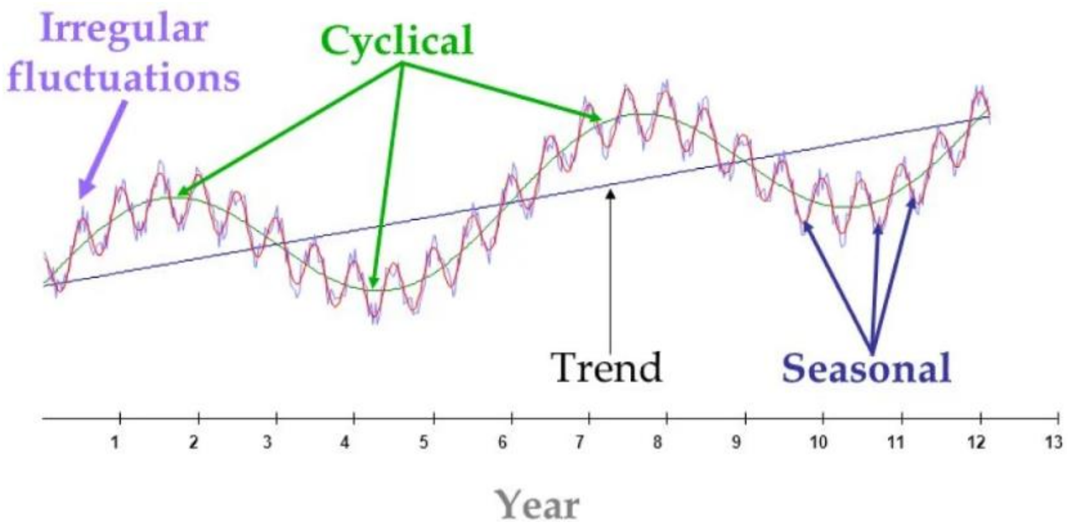
- Coffee price data is **highly nonlinear and volatile**
- One of the **most consumed soft commodity**

1. Weather / Climate	2. Seasonality / Cycles	3. Pests & Diseases	4. Agronomic Practices
5. Currency / Exchange Rates	6. Global Demand / Supply	7. Substitutes	8. Speculation & Futures Trading
9. Market Policies	10.Logistics / Supply Chain	11. Farmer Livelihood Assets	12. Labor Laws

Table 1: Factors Affecting Coffee Price

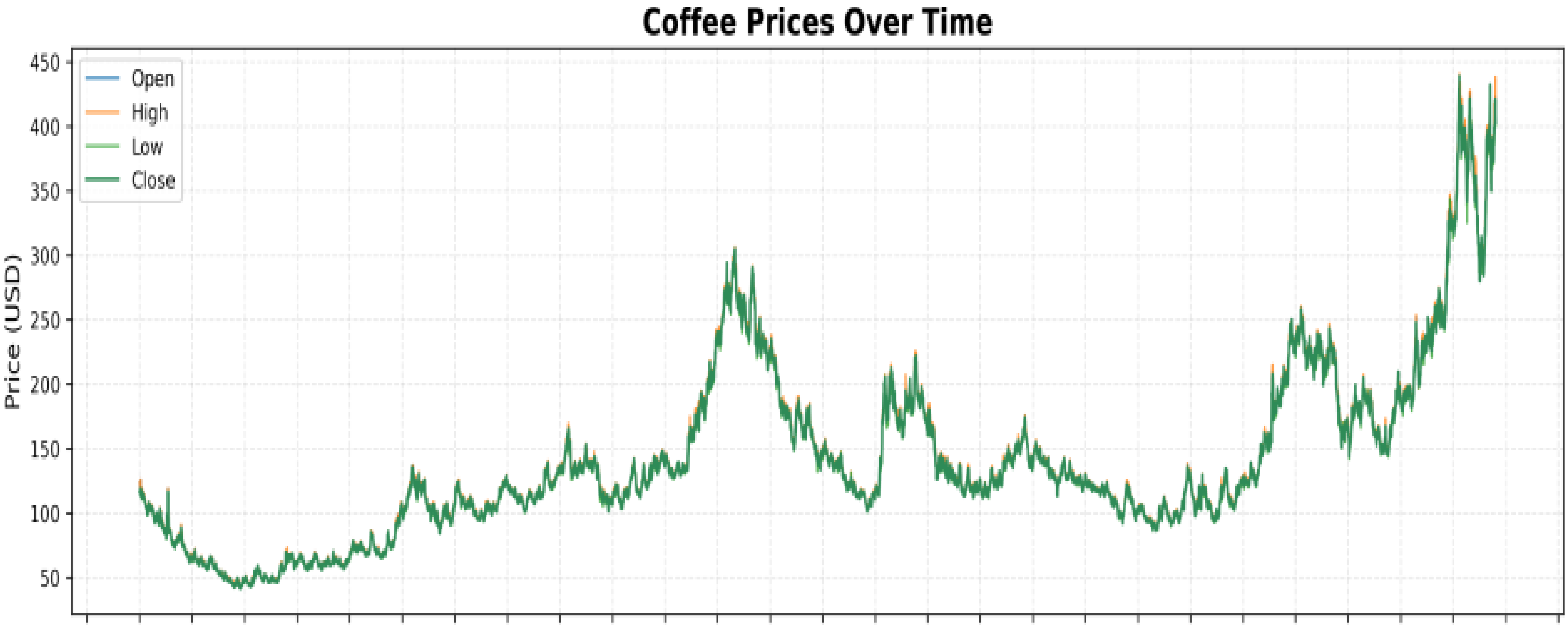
PREDICTION MODELS





Time Series Prediction

VISUALIZATION



TIME SERIES ANALYSIS

Understanding data behavior over time — a key step in accurate forecasting.

We checked for:

- **Stationarity** – Constant mean and variance over time.
- **Trend & Seasonality** – Long-term direction and repeating patterns.
- **Decomposition** – Splitting data into trend, seasonal, and residual parts.
- **Autocorrelation** – Relationship between current and past values.

TIME SERIES ANALYSIS - STATIONARITY

What It Means:

- A *stationary series* has a constant mean, variance, and autocorrelation over time.
- Many models (like ARIMA) assume data is stationary for stable forecasting performance.

ADF Test (Augmented Dickey-Fuller): Checks for a *unit root* — if present, the series is non-stationary.

KPSS Test (Kwiatkowski–Phillips–Schmidt–Shin): Checks for *trend-stationarity* — if the null is rejected, the series is non-stationary.

Our Findings (Coffee Prices):

ADF Test: $p = 0.987$ → Failed to reject non-stationarity.
KPSS Test: $p < 0.01$ → Rejected trend-stationarity.

➡ Conclusion: The *original series is non-stationary*.

TIME SERIES ANALYSIS- TREND AND SEASONALITY

What It Means:

- Trend represents the *long-term movement or direction* in prices.
- Seasonality reflects *repeating patterns or cycles* occurring at regular intervals.

Mann-Kendall Test: A non-parametric test used to identify the presence and direction of a trend (increasing or decreasing) in time-series data.It compares all pairs of observations to count how often later values are higher or lower than earlier

Our Findings :

Mann-Kendall Test: $p < 0.001$ → Significant *increasing trend*.

TIME SERIES ANALYSIS - AUTOCORRELATION

What It Means:

- Autocorrelation measures how current values in a time series are related to past values (lags).
- It helps identify repeating patterns, persistence, or randomness in the data.

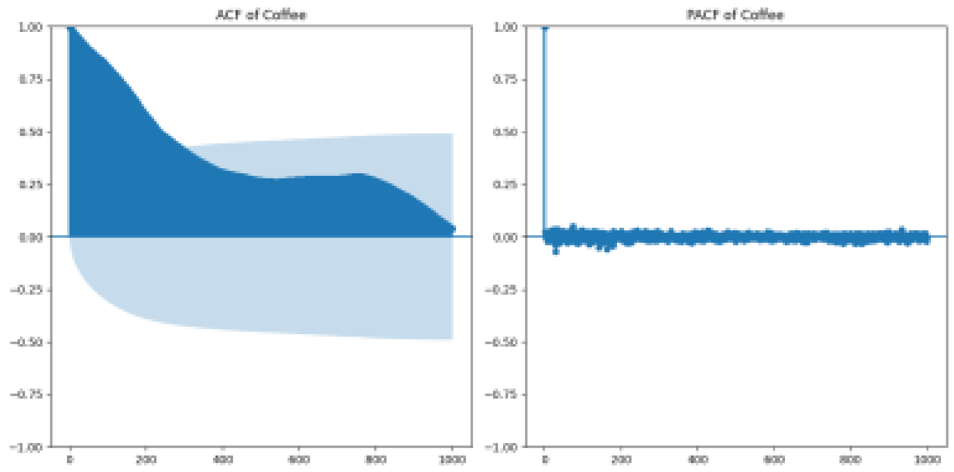
ACF (Autocorrelation Function): Quantifies how correlated a value is with its past values at different lags — reveals repeating patterns or persistence.

PACF (Partial Autocorrelation Function): Measures the correlation between a value and its lagged values after removing effects of shorter lags — useful for identifying direct dependencies.

Our Findings :

ACF Plot: Showed a *slow decay*, indicating a strong trend and long memory — a sign of non-stationarity.

PACF Plot: Revealed significant spikes at initial lags, suggesting short-term dependencies.



TIME SERIES ANALYSIS - DECOMPOSITION

What It Means:

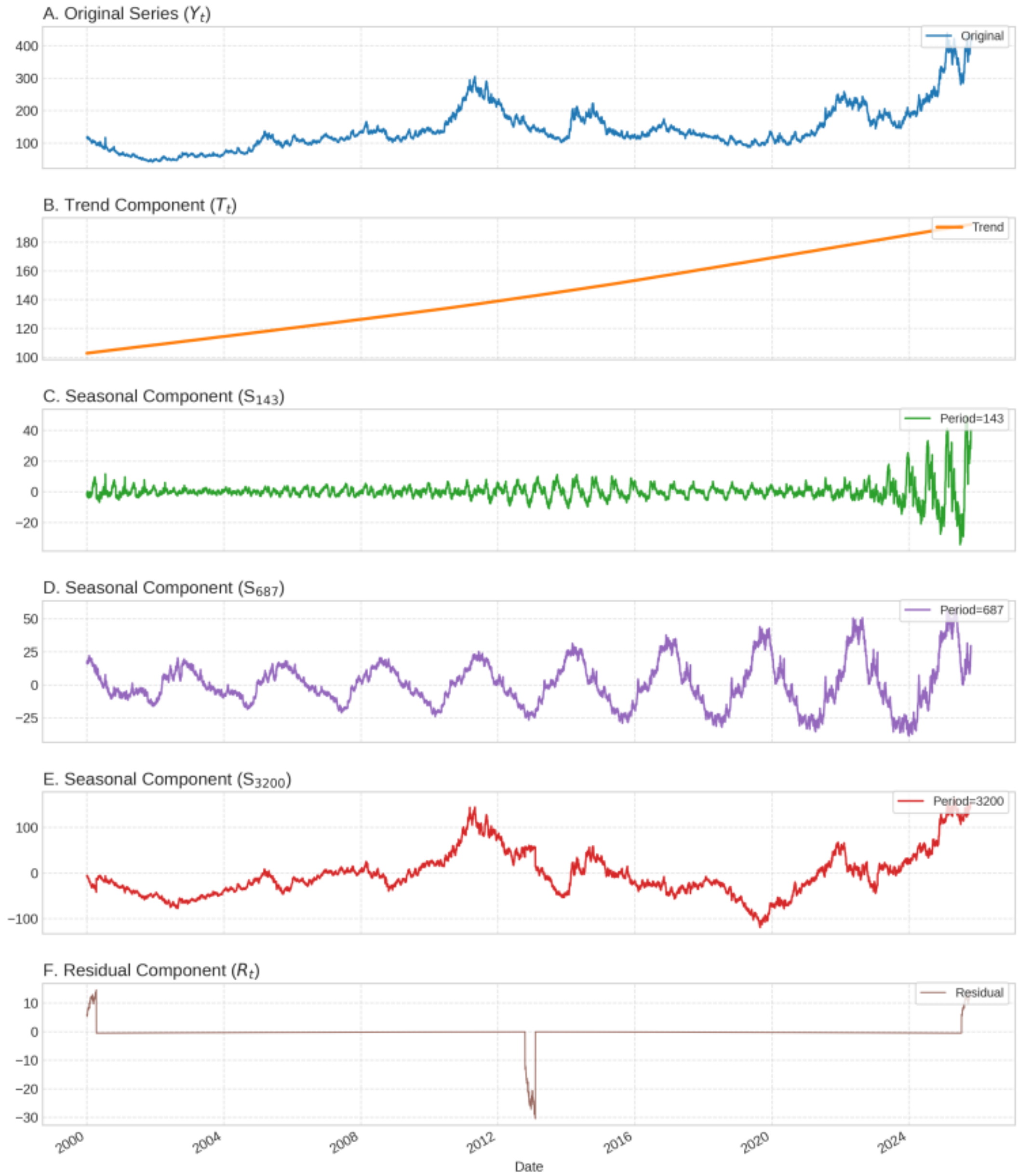
- Time series decomposition separates the data into:
  - Trend (T_t): Long-term direction of movement
  - Seasonal/Cyclical (S_t): Repeating or periodic patterns
  - Residual (R_t): Random or irregular component
- Helps reveal how much of the variation comes from trend, cycles, or randomness.

Multi-Seasonal Time Series Decomposition using Loess (MSTL)

Our Findings :

A brute-force search was performed using STL to identify possible seasonal periods using metrics such as the Ljung-Box test, seasonal strength, and trend strength across the original and differenced series. The results showed a strong trend but very weak seasonality, indicating that coffee prices lack consistent short-term seasonal patterns. with python cydets package we used MSTL decomposition with periods [143, 687, 3200 days].

TIME SERIES ANALYSIS - DECOMPOSITION



TIME SERIES ANALYSIS - SUMMARY

Category	Test Name	Statistic	p-value	Result/Inference
Stationarity	Augmented Dickey-Fuller (ADF)	0.568	0.987	Non-stationary Data
	KPSS	0.722	<0.01	Non-stationary Data
Trend & Seasonality	Mann-Kendall Test	(Z-stat) High pos.	<0.001	Significant increasing trend
	ACF-PACF plots	–	–	No repeating patterns or significant lag in the data
Decomposition	STL Decomposition	–	–	Failed to decompose successfully (on its own)
	Classical Additive / Multiplicative	–	–	Failed to decompose successfully
Autocorrelation	ACF/PACF Plots	–	–	Significant spikes but non-uniform lag

MOTIVATION

- Coffee price data shows a strong non-linear trend and multiple overlapping long-term cycles, making it too complex for a single model to capture effectively.

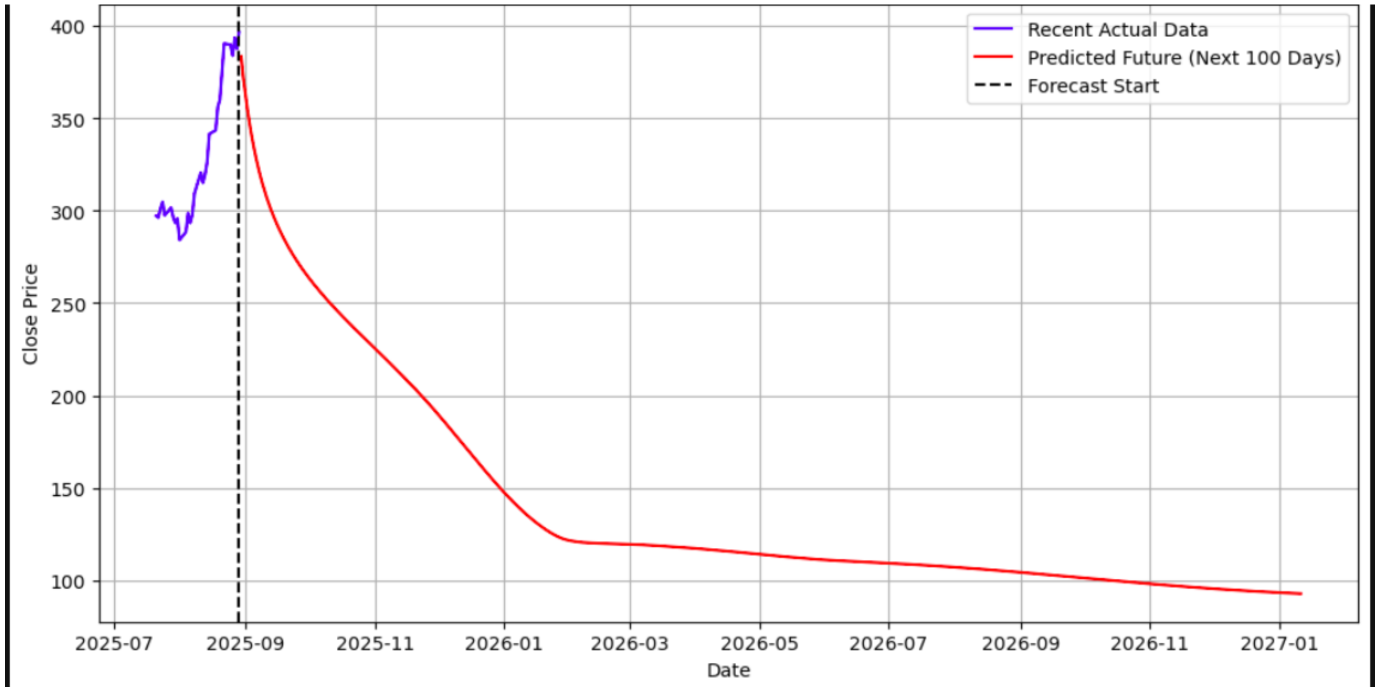
- A single LSTM tends to smooth out fluctuations, missing sharp cyclical movements critical to price volatility.

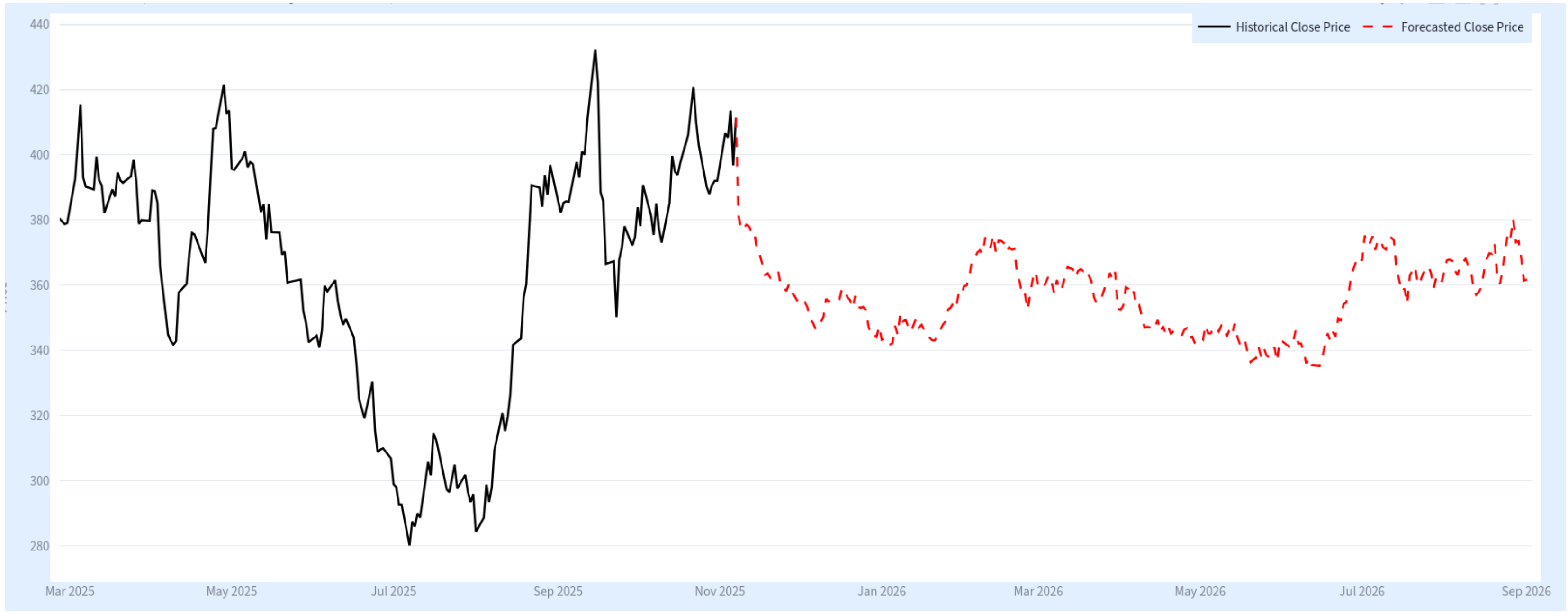
- To address this, a “divide and conquer” strategy was adopted — decomposing the series first and modeling each component separately.

- By forecasting the trend and cycles individually and recombining them, the final forecast becomes more realistic and volatility-aware than any single-model prediction.

MOTIVATION

Smoothened
forecast ->





Volatility Captured
forecast
<-

OUR HYPOTHESIS

Instead of modeling the complex series directly, the data is decomposed into simpler components — trend, cycles, and residuals.

🔹 Each component is modeled separately using the most suitable technique based on its behavior.

- Trend: smooth, long-term movement → stable regression model.
- Cycles & Residuals: irregular, non-linear → LSTM-based learning.

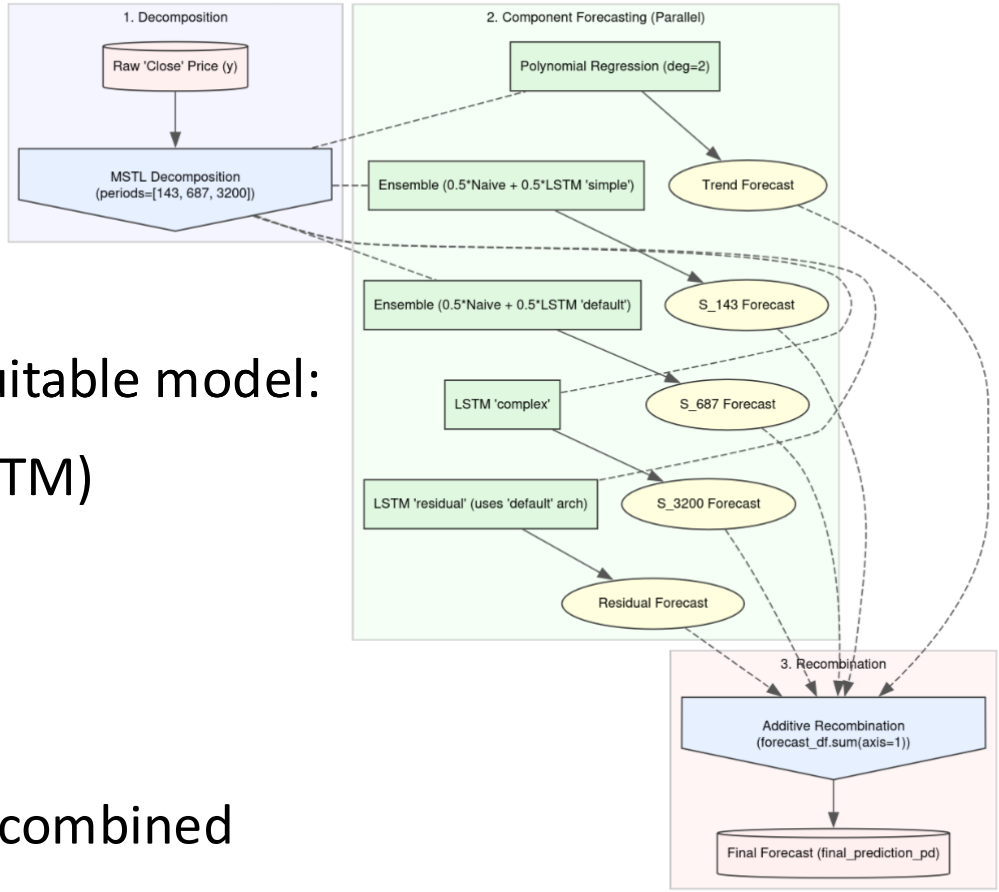
🔹 The individual forecasts are then recombined to form the final prediction.

🔹 This approach ensures the model captures both long-term direction and short-term volatility effectively.

MODEL DESIGN

The original series was split into five components:

- Trend (Tt)
- Cycle-1 (St,143)
- Cycle-2 (St,687)
- Cycle-3 (St,3200)
- Residual (Rt)



- Each component was forecasted independently using a suitable model:
  - Trend: Polynomial Regression (degree-2)
  - Cycle-1 & Cycle-2: 50/50 Weighted Ensemble (Naive + LSTM)
  - Cycle-3 & Residual: Dedicated LSTM models
- LSTM setup:
  - Lookback window = 60 days
  - Trained for 30 epochs, batch size = 32
  - Optimizer = Adam, Loss = MAE
- Final forecast: All component forecasts were additively recombined

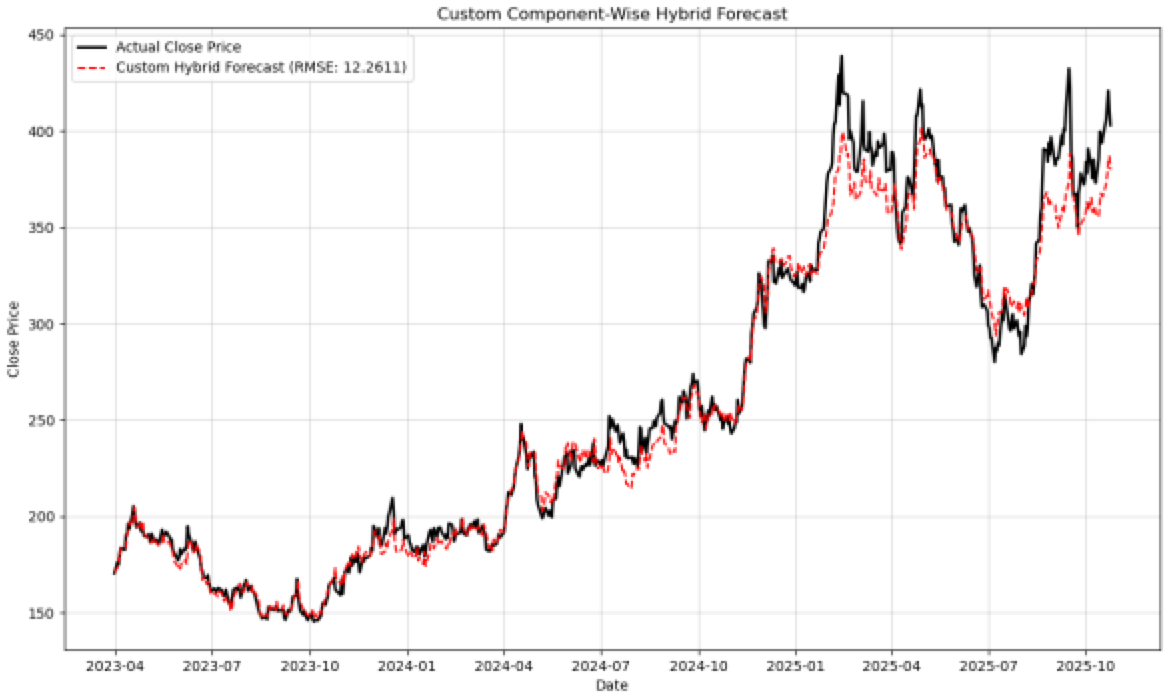
$$\hat{Y}_t = \hat{T}_t + \hat{S}_{t,143} + \hat{S}_{t,687} + \hat{S}_{t,3200} + \hat{R}_t$$

RESULTS

The hybrid model was evaluated on the test set (648 data points).

- Metrics used:
  - RMSE (Root Mean Squared Error)
  - MAE (Mean Absolute Error)
  - MAPE (Mean Absolute Percentage Error)

- Final Results:
  - RMSE = 12.26
  - MAE = 8.35
  - MAPE = 2.90%



STAKEHOLDERS INSIGHTS

- A 90-day forecast was analyzed to provide practical insights for different market participants.

- The forecast was divided into three time horizons:
  - Short-Term (1–7 days)
  - Medium-Term (30 days)
  - Long-Term (90 days)

- For each horizon, we examined:
  - Momentum (Trend): Direction of price movement (Upward / Downward / Stable)
  - Volatility: Standard deviation compared to recent history (High / Low)
  - Key Levels: Predicted peaks and troughs indicating expected trading range

- This analysis converts numerical forecasts into actionable signals for decision-making.

STAKEHOLDERS INSIGHTS

- Investor / Trader (Short-Term)
  - *Upward trend:* Consider long positions; watch volatility for stop placement.
  - *Downward trend:* Prefer short positions or put options; high volatility = higher risk.
  - *Stable trend:* Use range-trading or option-selling strategies.
- Retailer / Buyer (Medium-Term)
  - *Upward trend:* Secure inventory early; use fixed-price contracts.
  - *Downward trend:* Delay purchases or negotiate flexible pricing.
  - *Stable trend:* Continue regular procurement; focus on quality and efficiency.
- Producer / Farmer (Long-Term)
  - *Upward trend:* Hold inventory or delay sales to target higher prices.
  - *Downward trend:* Sell early or secure forward contracts.
  - *Stable trend:* Strengthen partnerships and ensure consistent buyers.

THANK YOU !

1. Artificial Intelligence in Finance: Coffee Commodity Trading Big Data for Informed Decision Making

Problems Prior Research:

1. **Incomplete data, documentation for storage**
2. **Requirement of scalable infrastructure**
3. **No system to store price data**

Results:

1. **Big-data Warehouse (Centralized Data Store)**
2. **Analytical Platform**
3. **Efficient data pipeline for prediction**

- 3 phase model: Data source, Data warehouse, Data consumption
- Data source : Web Crawling + Cleaning
- Data warehouse: Meta Data + Staging Data + ODS Data
- Data Mart: OLAP + statistical reporting + analysis
- Data consumption : ML Models
- Implementation : PostgreSQL + Python + Power BI + Prophet
- Factors Considered : Speculation / Future trading
- Data : Historical – **Yahoo Finance Portal + COT + CIT**
- Moving Average for signals.
- Data Warehousing is good but no analysis on which model is best to use? We can implement small scale Data Warehousing for project.

2. A methodology for Coffee Price Forecasting based on Extreme Learning Machine

Problems Prior Research:

1. **Improve the existing models to enhance accuracy**
2. **Accurate Non-linear model we not available**

Results:

1. ELM overcame the other models in 17 out of 24 scenarios (71%) considering all metrics. (considering MSE, MAE, MAPE).

- Used ELM as improvement over MLP to approximate values.
- Simplified training process + reduced network adjustment.
- Compared with ES, AR, ARIMA
- Works on stationary data (by differentiation)
- PACF for dominant lags and timeseries prediction
- Implementation : Minitab + Python
- Factors: Monthly average price (Historical Data)
- Data : Monthly Average prices of the ESALQ Coffee Indicator
- Model : ELM (Machine Learning)
- Done on monthly average.

3. A novel hybrid neural network-based volatility forecasting of agricultural commodity prices: empirical evidence from India

Problems Prior Research:

1. **Complex Non-linear Data of Agriculture.**
2. **ARCH and its improvisations capture volatility well but are linear.**
3. **LSTM cannot capture short term volatility but can capture non-linearity.**

Results:

1.Performed Better compared to GARCH, GJR-GARCH, MEM, EWMA. (considering RMSE, RNMSE, MAE, MAPE).

- Used a Hybrid : LSTM + GARCH
- LSTM captures non-linearity + GARCH volatility clustering
- LSTM captures, seasonal, cyclical and external causes
- GARCH can capture volatility clustering for short term data.
- LSTM predicts the next step's return -> GARCH applied on residual of predicted LSTM data.
- Implementation : Python software with 70% for training, 10% for validation and 20% for testing
- Data : 23 different commodities across 165 markets have been collected from the AGMARKNET
- GJR GARCH Can capture asymmetric shocks. Why not LSTM + GJR?

4. Improving trend prediction of agricultural futures price using image encoding and attention mechanisms

Problems Prior Research:

1. **Traditional econometric models rely on approximate linear assumptions**
2. **Machine learning still has limitations, such as sensitivity to parameters and a tendency to overfit.**

Results:

1. **Hybrid : ImgEnc-AttNet – CNN + LSTM + Attention Mechanism**
2. **Best Acc (Wheat) : ImgEnc-Att-RP - 68.92%**

- Integrating image encoding and attention mechanisms effectively harnesses rich latent information
- Sequence image generation : GAF, RP, Multichannel Images (temporal relations & spatial patterns)
- Image feature extraction : extract pertinent features from input market indicator images + channel attention
- Temporal feature extraction : Fuse image and raw numerical features
- Trend prediction : Final output.
- Implementation : Python + PyTorch
- Data : Chicago Mercantile Exchange

5. Meta-transformer: leveraging metaheuristic algorithms for agricultural commodity price forecasting

Problems Prior Research:

1. **Predetermined functional form**
2. **Models rely on stringent assumptions about data distribution.**
3. **Linear models (e.g., GARCH) miss nonlinear patterns**

Results:

1.Meta-Transformer (Transformer + MHAs) improved accuracy
2.70–90% better RMSE & MAE vs. GARCH/ANN/SVM/LSTM
3. WOA & GWO models gave best results

- Integrated MHASs: GWO, WOA, PSO
- Improved generalization, faster convergence, and enhanced predictive accuracy.
- Well suited for handling large datasets and enhancing forecasting accuracy.
- GWO and WOA exhibit simplicity, faster convergence, fewer parameters and robustness.
- Compared against statistical, ML, and DL benchmarks.
- Reduced forecasting errors → improved reliability.
- Dataset split: Train (90%), Validation (20% of train), Test (10%).

6. Enhancing agricultural commodity price forecasting with deep learning

Problems Prior Research:

1. **ARIMA & linear models fail with nonlinear, volatile data.**
2. **ML models (SVR, XGBoost) improve but miss long-term dependencies**
3. **Prior DL studies limited to small datasets/few commodities**

Results:

1. Deep learning (LSTM, GRU) outperformed ARIMA, SVR, XGBoost
2. GRU & LSTM achieved lowest RMSE, MAE, MAPE values

- Methods**:** Compared ARIMA, SVR, XGBoost, MLP, RNN, LSTM, GRU, ESN
- Better capture of nonlinear patterns & temporal dynamics, scalable, robust
- Implementation: Preprocessing (stationarity tests, normalization, sliding window), metrics (RMSE, MAE, MAPE, RNMSE)
- Dataset split: Train (80%) and Test (20%).

7. MACPGANA: design of a highly efficient multimodal agriculture commodity price prediction model via generative adversarial networks & autoencoders

Problems Prior Research:

1. **ARIMA & GARCH assume linearity → poor with nonlinear, volatile prices**
2. **Machine learning models lack temporal memory**
3. **LSTM/GRU capture sequences but need large data, risk overfitting**

Results:

1. **Proposed CNN–BiLSTM–Attention hybrid improved forecasting**
2. **Outperformed ARIMA, SVR, LSTM, GRU in RMSE & MAE**
3. **Achieved stable & accurate predictions for agri-price trends**

- Methods: CNN for feature extraction + BiLSTM for temporal learning + Attention for key dependencies
- Captures nonlinear + long-term patterns, reduces error, interpretable
- Data: Daily wholesale price data of multiple agricultural commodities (AGMARKNET, 2010–2024)
- Implementation: Preprocessing (smoothing, normalization), metrics (RMSE, MAE, MAPE)

Plan:

- Data : Yahoo Finance Portal
- Small scale data warehousing (if feasible)
- LSTM + GARCH **OR** Transformer + WOA ?

- **We are interested in LSTM based Models for prediction:**
- 1.LSTM
- 2.BiLSTM
- 3.LSTM+GARCH
- 4.BiLSTM + GARCH
- 5.LSTM + GJR-GARCH
- 6.BiLSTM + GJR-GARCH

References:

1. Ngoc-Bao-Van Le, Yeong-Seok Seo, and Jun-Ho Huh (2024). **Artificial Intelligence in Finance: Coffee Commodity Trading Big Data for Informed Decision Making**. IEEE Access. doi: 10.1109/ACCESS.2024.3409762

2. C. Deina et al. (2021**). A methodology for coffee price forecasting based on extreme learning machines. Information Processing in Agriculture**. doi: 10.1016/j.inpa.2021.07.003

3. R. L. Manogna, Vijay Dharmaji, and S. Sarang (2025). **A novel hybrid neural network-based volatility forecasting of agricultural commodity prices: empirical evidence from India**. Journal of Big Data. doi: 10.1186/s40537-025-01131-8

4. Dabin Zhang et al. (2025**). Improving trend prediction of agricultural futures price using image encoding and attention mechanisms. Management System Engineering**, 4, 10. doi: 10.1007/s44176-025-00042-5

5. G. H. Harish Nayak et al. (2025). **Meta-transformer: leveraging metaheuristic algorithms for agricultural commodity price forecasting. Journal of Big Data**. doi: 10.1186/s40537-025-01196-5

6. R. L. Manogna, Vijay Dharmaji, and S. Sarang (2025**). Enhancing agricultural commodity price forecasting with deep learning.** Scientific Reports, 15, 20903. doi: 10.1038/s41598-025-05103-z

7. Rubi Kambo et al. (2025). **MACPGANA: design of a highly efficient multimodal agriculture commodity price prediction model via generative adversarial networks & autoencoders**. International Journal of Information Technology. doi: 10.1007/s41870-025-02425-z