PREDICTING NATURAL RUBBER PRICES:
AN ANALYSIS OF ARIMA-GARCH, EXPONENTIAL SMOOTHING AND LSTM APPROACHES

Presented by Liau Ling Yee & Pun Yi Jie
Supervised by Dr. Ong Wen Eng

OUTLINE

- INTRODUCTION
- PROBLEM STATEMENT
- OBJECTIVES
- LITERATURE REVIEW
- METHODOLOGY
- RESULT & DISCUSSION
- CONCLUSION
- FUTURE WORK

INTRODUCTION

INTRODUCTION

Natural Rubber

- An integral component of Malaysia's economy.
- World’s leading producer and exporters of multifunctional materials.

Malaysia's Natural Rubber Industry

- A major player in the global market due to its favorable climate and expertise.

Economic and Social Importance

- Supports livelihoods of smallholders, promotes rural development, and is a key part of Malaysia's agricultural heritage.

Malaysia's Position in Global Trade

- Malaysia's ranking as 5th largest exporter (ANRPC report, 2022) with China as the main export destination.

SMR20 Rubber

- A key export product with valuable properties for tires and industrial products.

INTRODUCTION

What is autoregressive integrated moving average – generalized autoregressive conditional heteroscedasticity (ARIMA-GARCH)?

- Hybrid Statistical Approach
- ARIMA - trend and seasonality
- GARCH - volatility

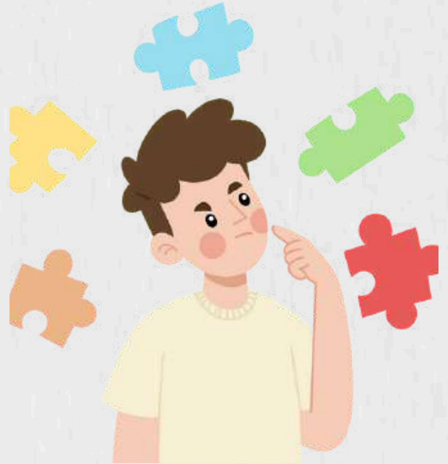
What is exponential smoothing (ES)?

- Statistical Approach
- Assigns weights to past observation, recent data has higher weight.

What is long short-term memory (LSTM)?

- Machine Learning Approach
- Designed to handle long-term dependencies in sequential data

PROBLEM STATEMENT



PROBLEM STATEMENT

Global Leader in Rubber Products

- World's largest supplier of medical gloves, catheters, etc.

SMR20 Price Volatility

- Fluctuates due to global factors, weather, and political events.

Importance of Price Prediction

- Crucial for informed decisions by farmers, exporters, and policymakers.

Forecasting Methods

- Statistical (ARIMA-GARCH, Exponential Smoothing) vs. Machine Learning (LSTM networks).

OBJECTIVES

OBJECTIVES

- To investigate the price prediction of natural rubber SMR20 in the literature.
- To study the usage of autoregressive integrated moving average – generalized autoregressive conditional heteroscedasticity (ARIMA-GARCH), exponential smoothing (ES), and long short-term memory (LSTM) in SMR20 rubber price prediction.
- To implement ARIMA-GARCH, ES and LSTM in Python for SMR20 rubber price prediction and compare the performance of the models.

LITERATURE REVIEW

LITERATURE REVIEW

Articles	Method	Work	Conclusion
Fu & Jamaludin, 2022	Autoregressive Integrated Moving Average (ARIMA)	Bulk latex price forecasting	• **ARIMA effectively predicts** bulk latex prices • A **lower MAPE** of 8.59 percent and **RMSE** of **69.78 sen per kilogram**
Dritsaki, 2018	Autoregressive Integrated Moving Average - Generalized Autoregressive Conditional Heteroscedasticity (ARIMA-GARCH)	Oil price forecasting	• **ARIMA-GARCH** outperforms traditional ARIMA by better handling **volatility** and **nonlinearity**

LITERATURE REVIEW

Articles	Method	Work	Conclusion
Fatima *et al.*, 2019	Exponential Smoothing (ES)	Carbon dioxide emission forecasting	• **Simple exponential smoothing** is best suited for Pakistan and Sri Lanka based on **minimum FMAE** • Selection of forecasting model should be tailored to **specific data characteristics**
Khairina *et al.*, 2021		Local water company income forecasting	• **Double exponential smoothing** outperforms triple exponential smoothing • Achieved a **MAPE** of **9.54%**, demonstrating **higher accuracy**

LITERATURE REVIEW

Articles	Method	Work	Conclusion
Chen *et al.*, 2017	Long Short-Term Memory (LSTM)	House price forecasting	• **LSTM** shows excellent properties and noticeable **improvement** in accuracy compared to the baseline ARIMA model • The results of stateful LSTM and stacked LSTM models are not significantly better than the basic LSTM model
Yildirim *et al.*, 2023		Electricity market price forecasting	• **LSTM** could generate **reliable forecasts**, effectively capturing trends and patterns in LMP changes, even under **significant disruptions**

METHODOLOGY

MAIN FLOW

Data Preparation → Exploratory Data Analysis (EDA) → Data Splitting → Analysis

Exploratory Data Analysis (EDA):

- Descriptive statistics
- Data visualization

Data Splitting:

- Train set
- Test set

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE - GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY (ARIMA-GARCH)

STEP 1:

Test for stationarity using ADF test

→

STEP 2:

Fit ARIMA model by determining p, d, q

p: no. of autoregressive terms
d: no. of differences needed
q: no. of lagged forecast errors

→

STEP 3:

Diagnose residuals and check assumption

→

STEP 4:

Apply GARCH to residuals to solve heteroscedasticity

EXPONENTIAL SMOOTHING (ES)

STEP 1: Model selection → STEP 2: Fit double exponential smoothing model → STEP 3: Forecasting

LONG SHORT-TERM MEMORY (LSTM)

STEP 1:

Find optimal sequence length

→

STEP 2:

Hyperparameter Tuning

- Batch Size
- Dropout Rate
- Epochs
- Learning Rate
- Units

→

STEP 3:

Fit LSTM model using the best combination of parameters

→

STEP 4:

Forecasting

RESULT & DISCUSSION

ARIMA-GARCH MODEL

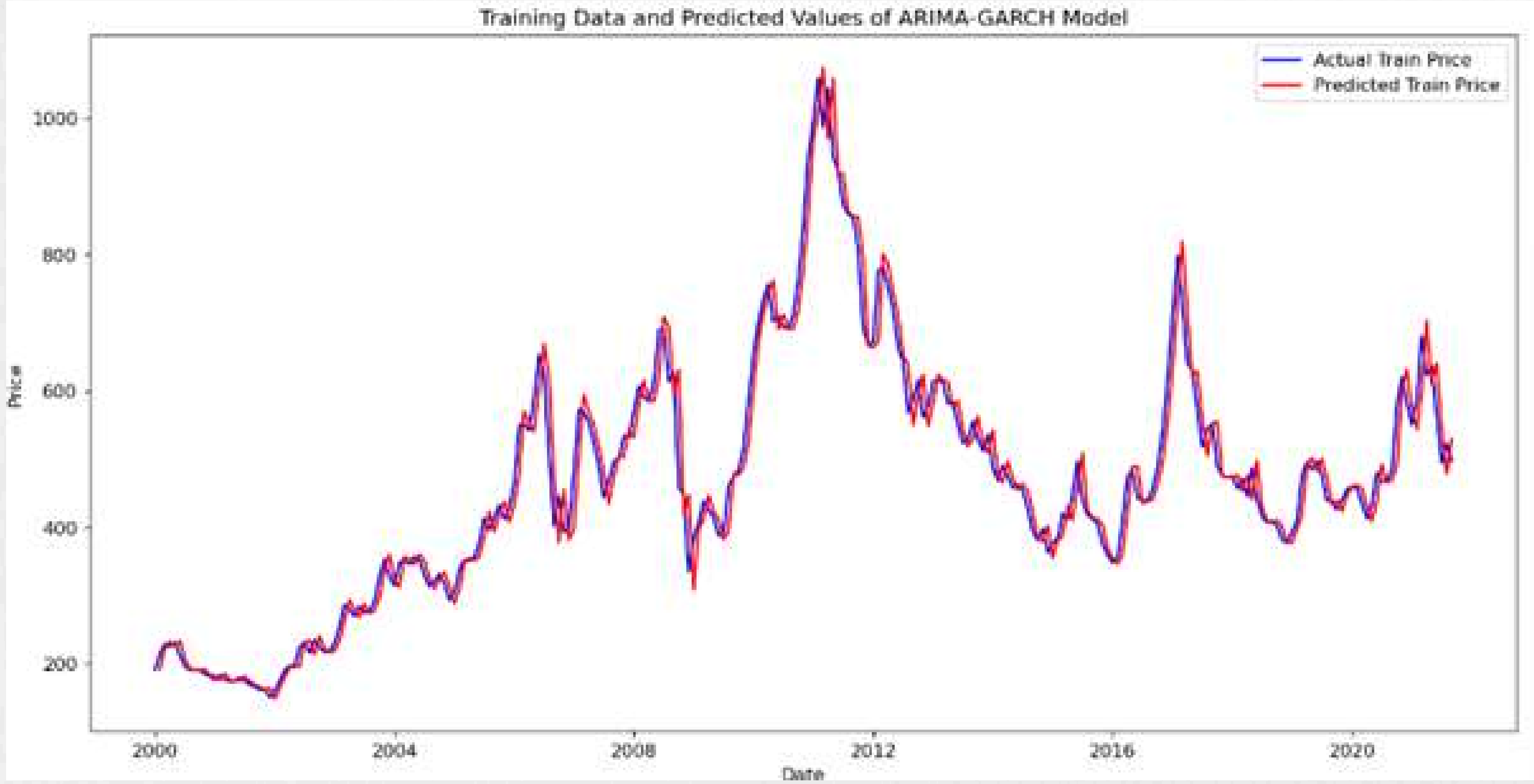
```
Best model:  ARIMA(1,1,0)(0,0,0)[0]
Total fit time: 1.008 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  261
Model:               SARIMAX(1, 1, 0)   Log Likelihood               -1316.226
Date:                Tue, 18 Jun 2024   AIC                           2636.452
Time:                        22:51:44   BIC                           2643.573
Sample:                    01-01-2000   HQIC                          2639.315
                         - 09-01-2021
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.2361      0.048      4.928      0.000       0.142       0.330
sigma2      1460.7659     91.566     15.953      0.000    1281.300    1640.232
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                56.11
Prob(Q):                              0.90   Prob(JB):                         0.00
Heteroskedasticity (H):               1.68   Skew:                            -0.57
Prob(H) (two-sided):                  0.02   Kurtosis:                         4.97
===================================================================================
```

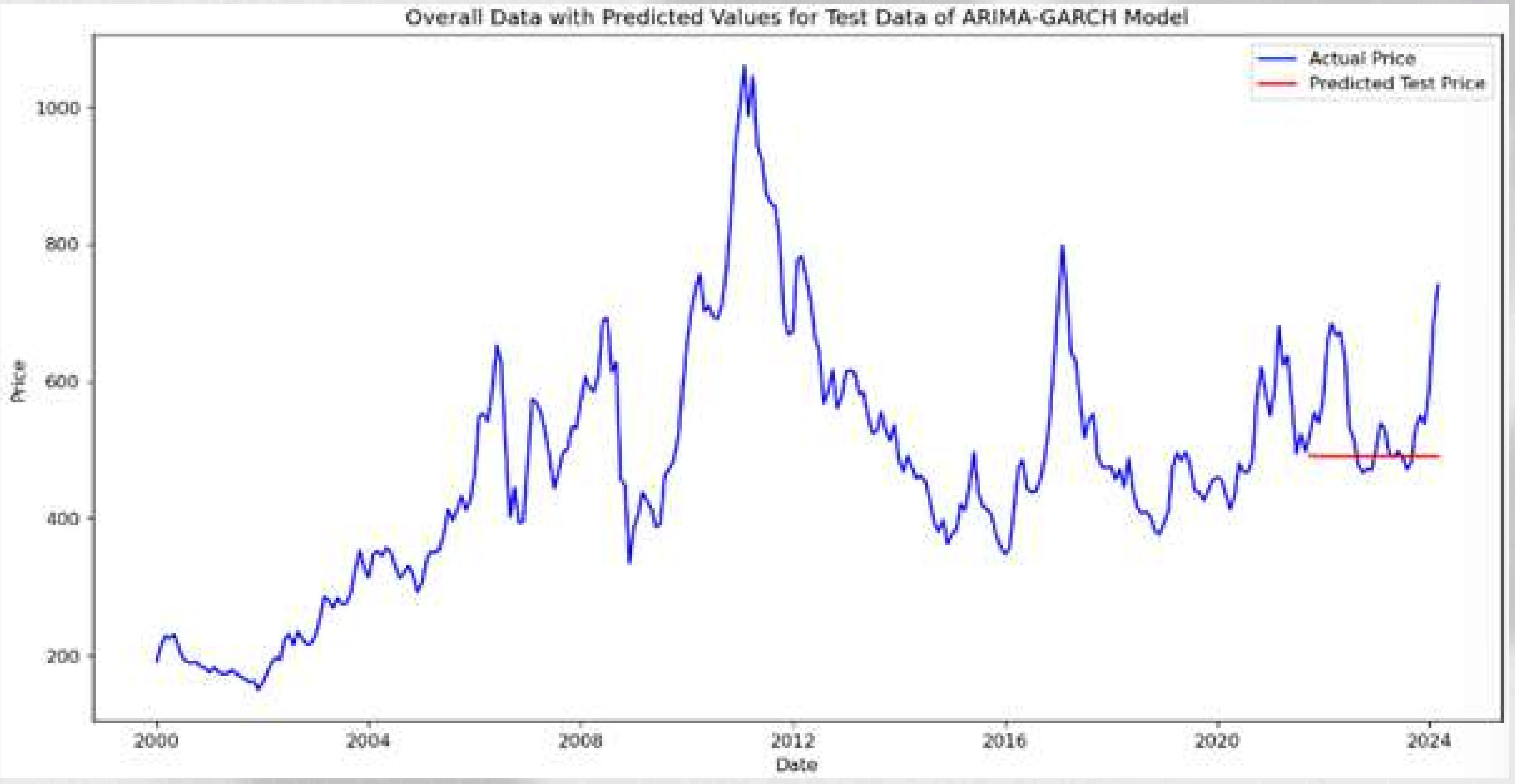
- ARIMA(1,1,0) model fitted to differenced series
- Significant coefficients and residuals indicate good fit

```
                     Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                      0   R-squared:                       0.000
Mean Model:             Constant Mean   Adj. R-squared:                  0.000
Vol Model:                      GARCH   Log-Likelihood:               -1294.71
Distribution:                  Normal   AIC:                           2597.42
Method:            Maximum Likelihood   BIC:                           2611.68
                                        No. Observations:                  261
Date:                Tue, Jun 18 2024   Df Residuals:                      260
Time:                        22:51:45   Df Model:                            1
                               Mean Model
============================================================================
                 coef    std err          t      P>|t|   95.0% Conf. Int.
----------------------------------------------------------------------------
mu             0.8330      1.729      0.482      0.630 [ -2.555,  4.221]
                              Volatility Model
==============================================================================
                 coef    std err          t      P>|t|     95.0% Conf. Int.
------------------------------------------------------------------------------
omega         71.7837     65.729      1.092      0.275 [-57.042,2.006e+02]
alpha[1]       0.3950      0.129      3.070  2.140e-03   [  0.143,  0.647]
beta[1]        0.6050  9.481e-02      6.381  1.754e-10   [  0.419,  0.791]
==============================================================================
```

- GARCH(1,1) Model
- Solves heteroscedasticity in residuals
- Significant ARCH and GARCH terms

ARIMA-GARCH MODEL PERFORMANCE





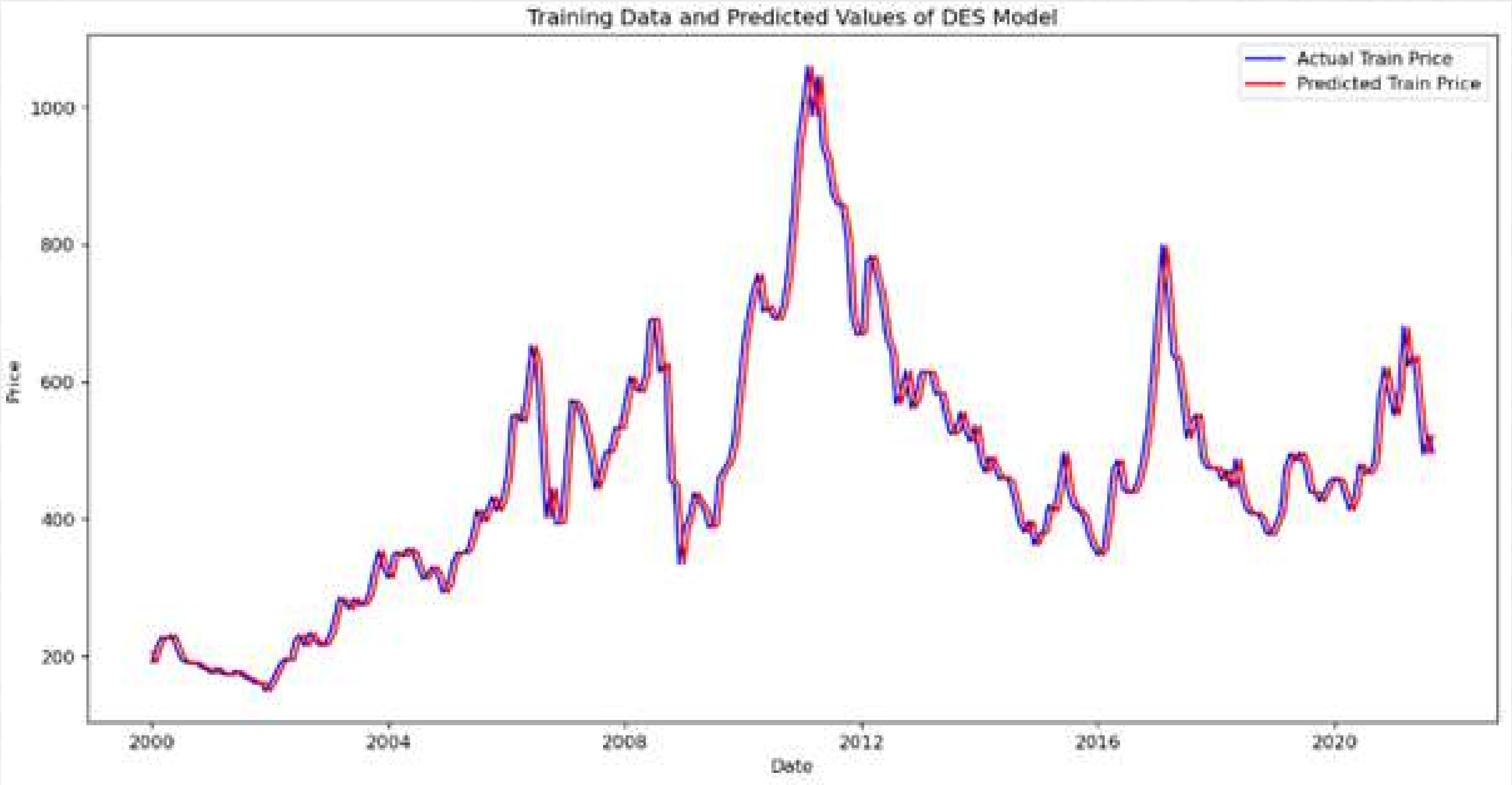
Training Data vs Predicted Values

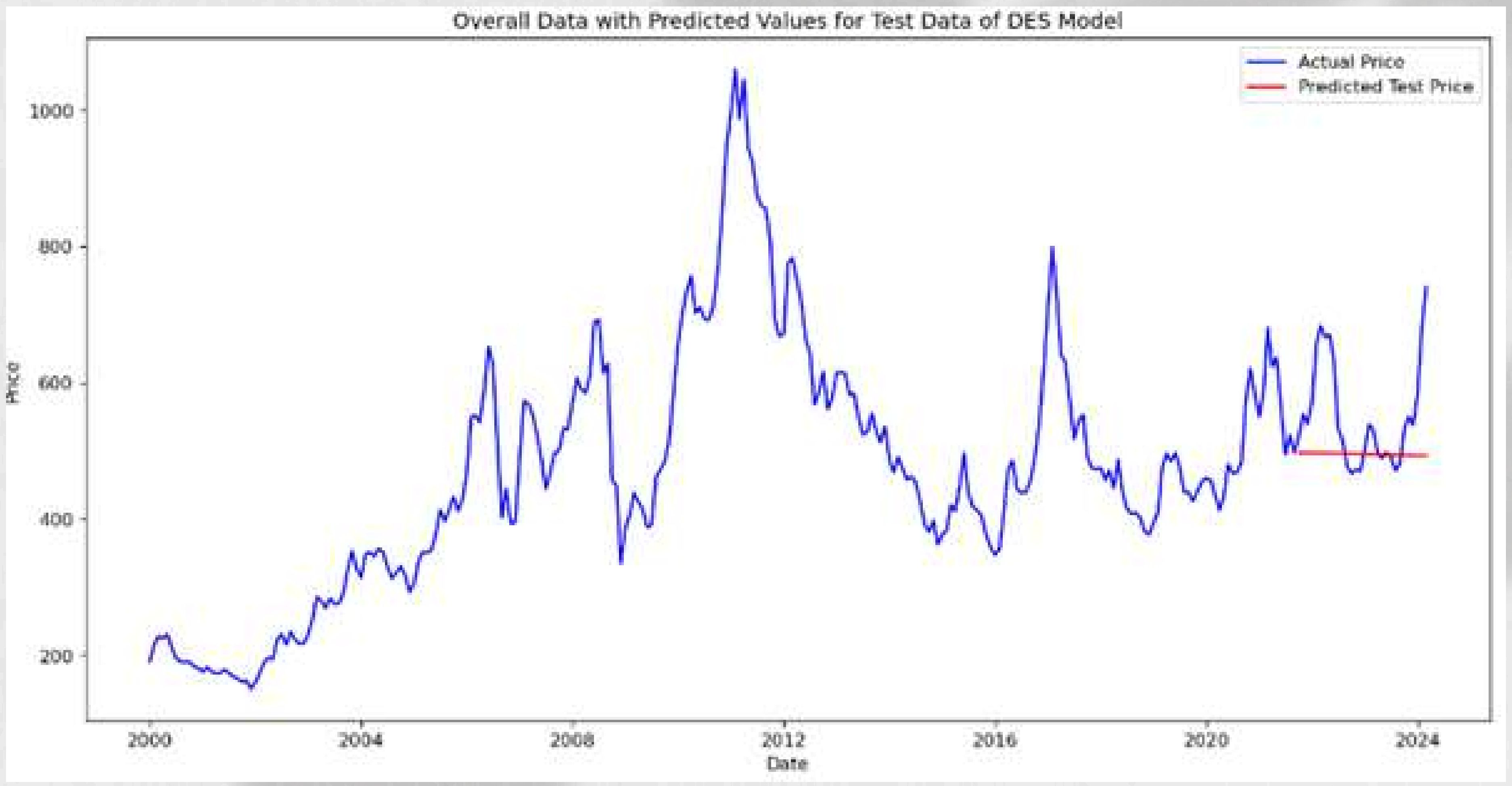
- Predicted prices behave similarly to actual prices
- Accurately represents the general trend and volatility
- Expected values are close to actual prices

Overall Data vs Test Data Predictions

- Predicted test prices indicate a steady trend with less volatility
- Does not reflect the extreme variations and peaks observed in the actual data

EXPONENTIAL SMOOTHING MODEL PERFORMANCE





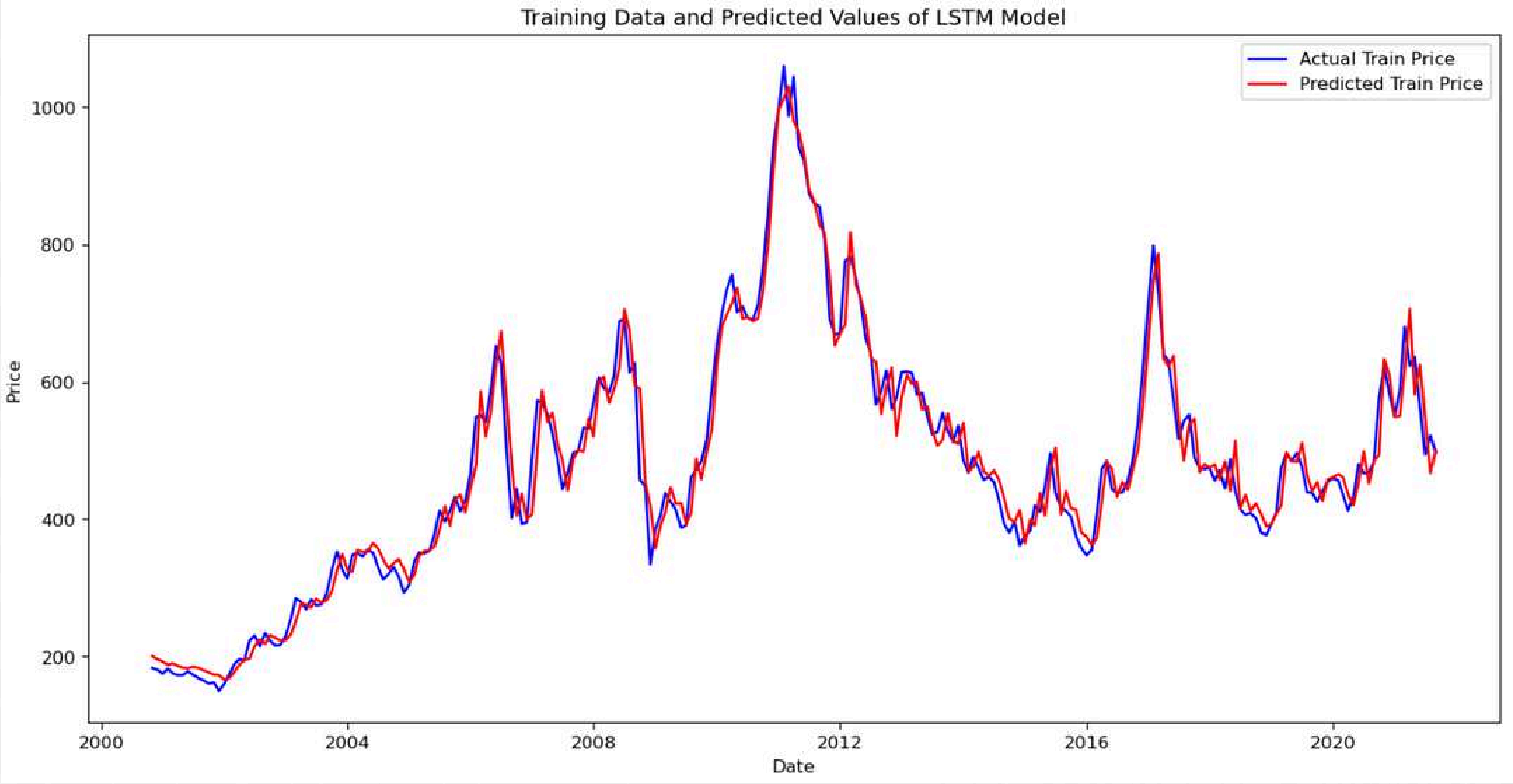
Training Data vs Predicted Values

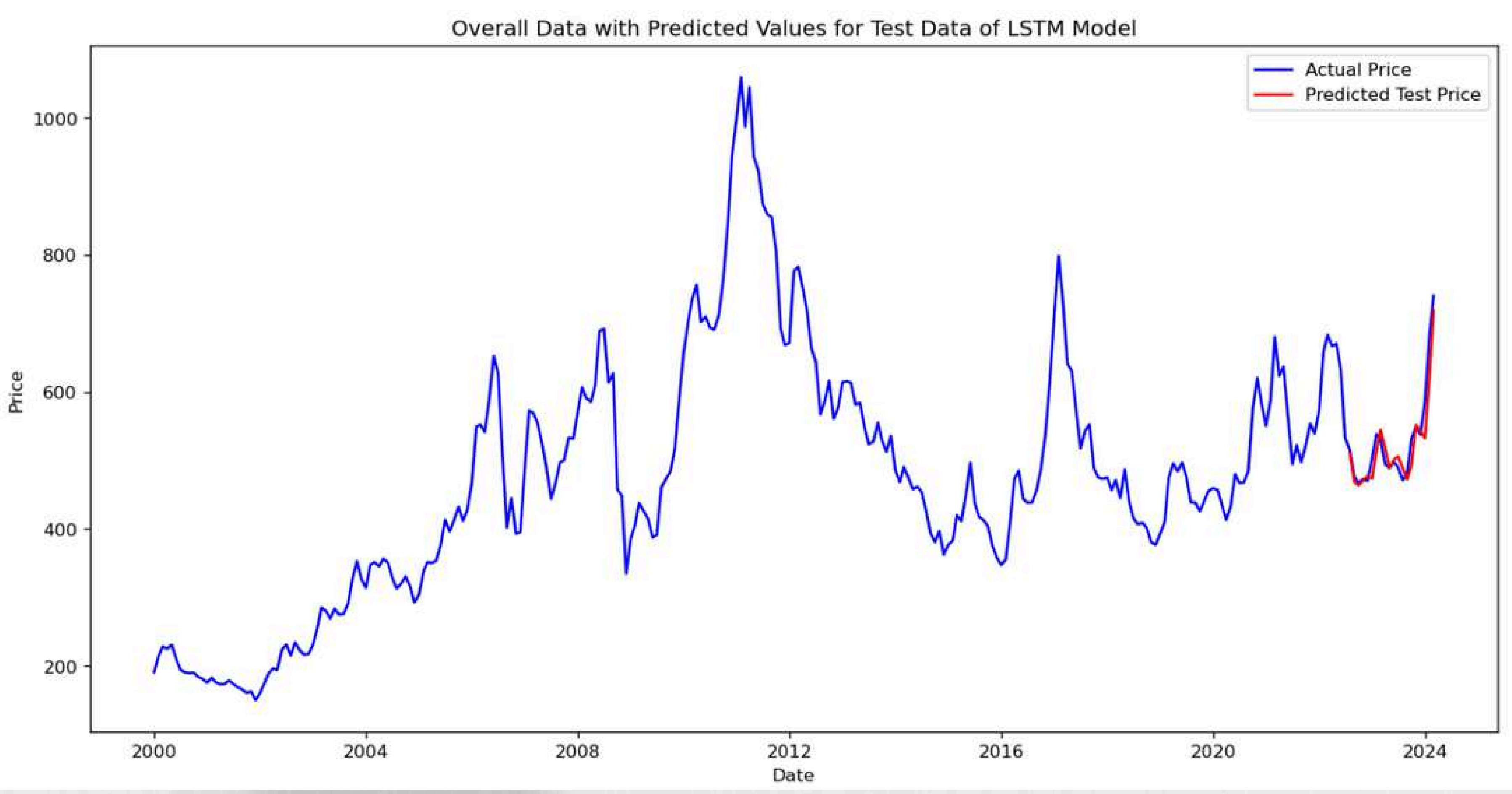
- Strong fit
- Captures major trends

Overall Data vs Test Data Predictions

- Fails to capture significant variations

LSTM MODEL PERFORMANCE





Training Data vs Predicted Values

- Predicted prices behave similarly to actual prices
- Accurately represents the general trend and volatility

Overall Data vs Test Data Predictions

- Predicted values appear to generally follow the upward trend of the actual prices.
- Closely alignment of the two lines suggests that the LSTM model successfully capture the dependencies among the observation in the series.

PERFORMANCE COMPARISON

Model	RMSE	MAE	MAPE	R^2
ARIMA-GARCH	98.0041	68.4460	11.0615	-0.6466
Exponential Smoothing	95.0886	66.1419	10.6995	-0.5500
LSTM	24.5575	17.4887	3.1496	0.8764

- LSTM has the lowest RMSE, a smaller difference between predicted and actual values.
- LSTM has the lowest MAE and MAPE, highlighting the ability to generate forecasts closer to the actual observation.
- LSTM has the highest R-squared, signifying a strong positive fit between predicted and actual values.
- Negative R-squared for ARIMA-GARCH and exponential smoothing, fails to explain a significant portion of the variance in the actual data.

CONCLUSION

CONCLUSION

Unexpected outcome

- ARIMA-GARCH model was used to capture both linear patterns and volatility clustering in the series
- Exponential smoothing was applied since it is simple but effective in treating non-seasonal data.
- Both obtained poor forecasting result.

Best Model

- LSTM model significantly outperformed the ARIMA-GARCH and exponential smoothing models across all evaluation metrics.
- Smaller difference between predicted and actual values.
- Large R-squared value.

FUTURE WORK

DIRECTIONS FOR FUTURE RESEARCH

Incorporate Additional Variables

- Rubber Production Volume
- Rubber Export Volume
- Local Processing Costs

Explore Alternative Volatility Models

- EGARCH
- TGARCH

Optimize Exponential Smoothing Parameters

- Grid Search
- Cross-Validation
- Simulated Annealing

Hyperparameter Optimization for LSTM

- Simulated Annealing
- Genetic Algorithms
- Particle Swarm Optimization
- Symbiotic Organism Search

REFERENCES

- Chen, X., Wei, L., & Xu, J. (2017). House price prediction using LSTM. arXiv preprint arXiv:1709.08432.
- Ghani, I. M. M., & Rahim, H. A. (2024). Building a Sustainable GARCH Model to Forecast Rubber Price: Modified Huber Weighting Function Approach. Baghdad Science Journal 21(2), 0511-0511.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation 9 (8), 1735-1780. https://ieeexplore.ieee.org/abstract/document/6795963
- Jie, L. C., & Lee, M. H. (2022). Autoregressive Integrated Moving Average, Exponential Smoothing and Artificial Neural Network to Forecast Natural Rubber Prices in Malaysia.
- Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing 323, 203-213. https://doi.org/10.1016/j.neucom.2018.09.082

THANK YOU