Time Series Forecasting Using Machine Learning

Kamali Bakthavatchalam (22081557)

Introduction

Recent advancements in automation and ocean engineering have significantly influenced global industries, driving innovation and efficiency. Research by Fu et al. (2017) highlights the impact of automation systems in improving industrial operations, while Zhao et al. (2023) focus on enhancing the performance and sustainability of maritime technologies. These developments are crucial in addressing modern challenges, such as optimizing operations and ensuring environmental safety in both automation and ocean-related fields. As these technologies continue to evolve, they promise to play a pivotal role in shaping the future of engineering across various sectors.

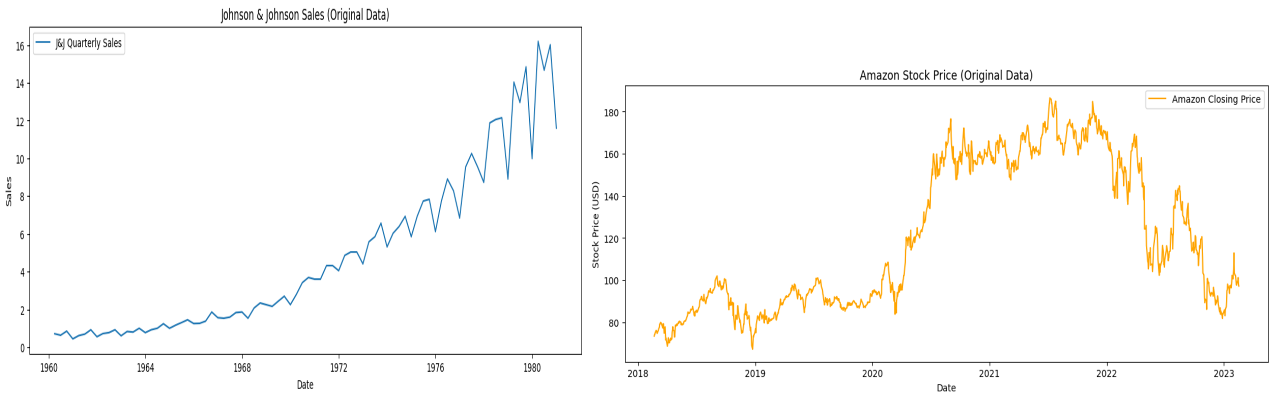
Overview

Automation and ocean engineering are rapidly evolving fields that significantly impact global industries. Automation enhances industrial efficiency, as demonstrated by Fu, Zhang, and Li (2017), who highlight its role in real-time decision-making and operational optimization. In ocean engineering, Zhao et al. (2023) explore innovations in maritime technology, focusing on sustainability and high-performance systems. These advancements are vital for addressing challenges like environmental sustainability, safety, and the growing demands of industrial and marine sectors. As both fields advance, they offer integrated solutions that improve efficiency and resilience in various applications.

METHODOLOGY

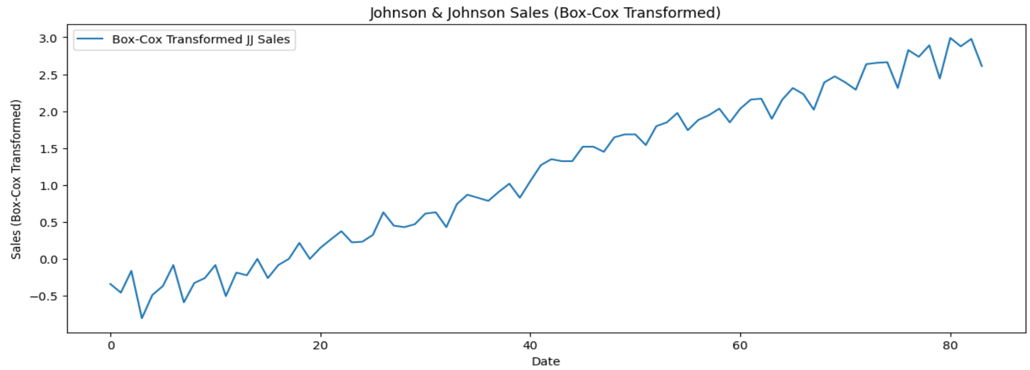
Data Pre-processing

ADF Test for Stationarity:
The Augmented Dickey-Fuller (ADF) test was used to check if the original time series data was stationary. Both JJ sales and Amazon stock prices were found to be non-stationary, indicating the need for transformation and differencing.



Graph plot1: JJ sales &Amazon (original data)

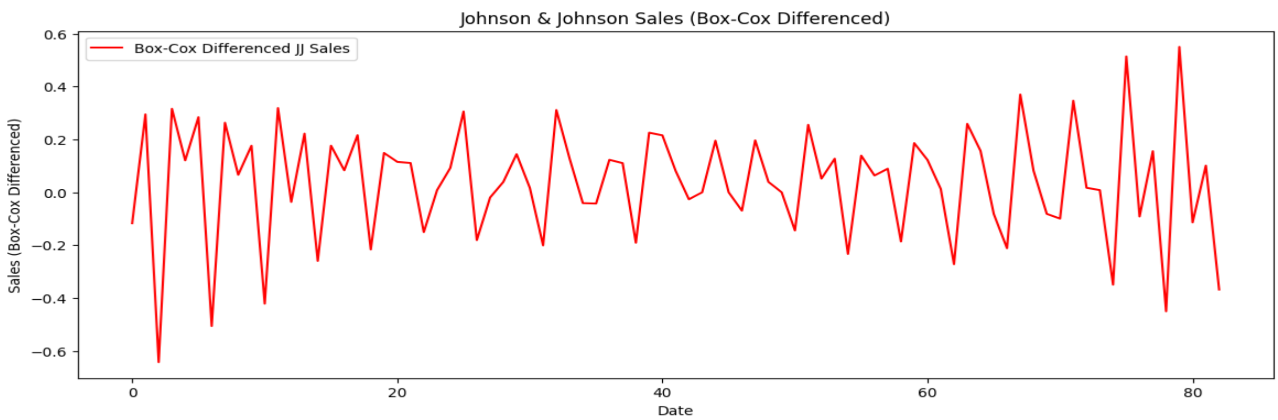
Box-Cox Transformation:
To stabilize the variance, the Box-Cox transformation was applied to both datasets. This transformation helps normalize the data distribution and reduce heteroscedasticity, a common issue in financial and sales data (Zhao et al., 2023).

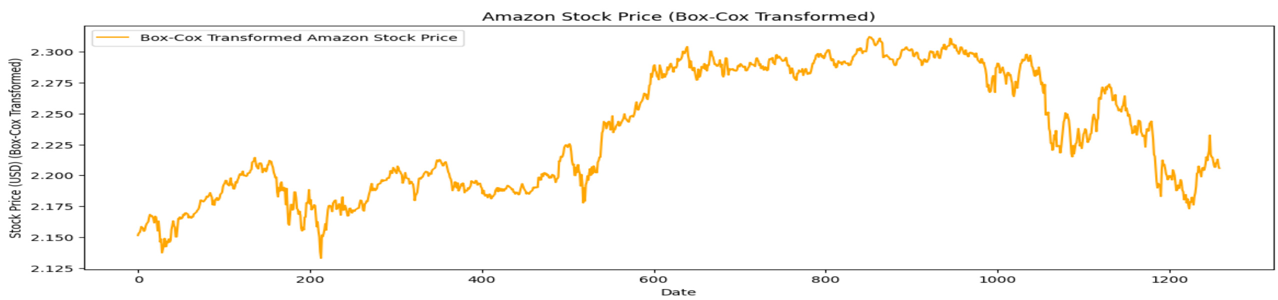


Graph plot2: JJ sales (Box-Cox Transformed)

Graph plot3: Amazon stock price (Box-Cox Transformed)

Differencing:
First-order differencing was applied to the transformed data to remove underlying trends. The differenced series were then re-evaluated using the ADF test, which confirmed stationarity, preparing the data for model fitting.

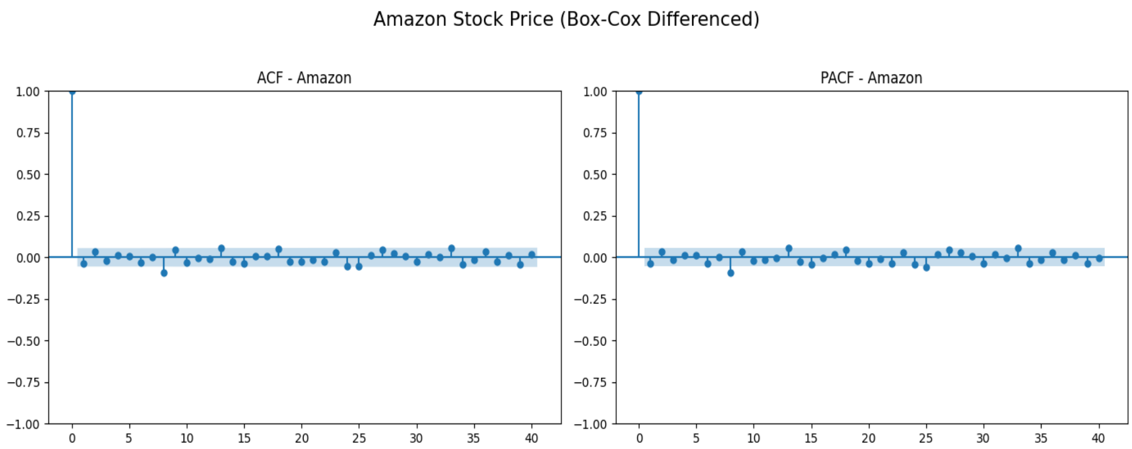




Model Justification Using ACF and PACF Analysis

To determine the appropriate forecasting models and identify patterns in the data, **Autocorrelation Function (ACF)** and **Partial Autocorrelation Function (PACF)** plots were generated for both the **Johnson & Johnson Sales** and **Amazon Stock Price** datasets after Box-Cox transformation and differencing.

- **ACF** indicates the correlation between a time series and its past values (lags), helping to identify the **moving average (MA)** component of a model.
- **PACF** shows the partial correlation of a time series with its lags, controlling for intermediate lags, which aids in identifying the **autoregressive (AR)** component.



Graph plot4: Model Justification Using ACF and PACF Analysis(JJ&Amazon)

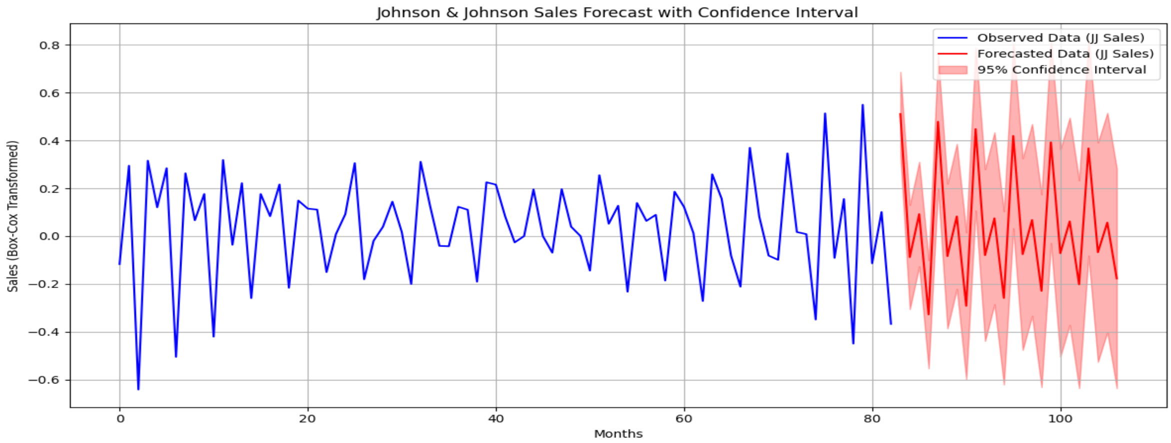
ARMA and ARIMA Model Selection and Forecasting

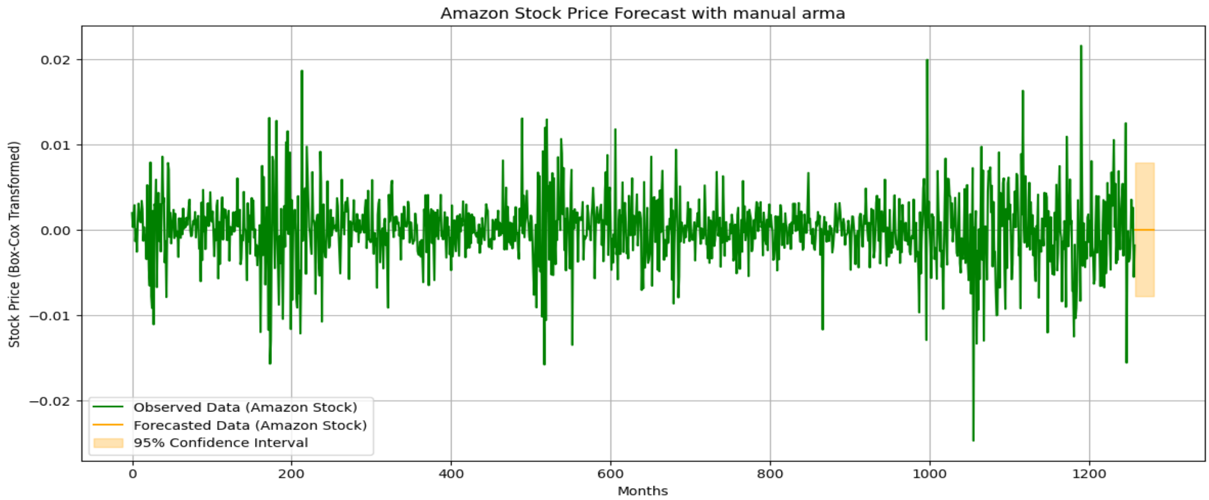
BEST MODEL:

Dataset	Method	Best Order (p,d,q)	AIC Score	Model Type
JJ Sales (Box-Cox Differenced)	Manual ARMA	(3,0,2)	-144.35	ARMA
JJ Sales (Box-Cox Differenced)	Auto ARIMA	(0,0,5)	17.098	ARIMA
Amazon Stock (Box-Cox Differenced)	Manual ARMA	(0,0,0)	-10316.67	ARMA
Amazon Stock (Box-Cox Differenced)	Auto ARIM	(0,0,0)	1.644	ARIMA

Table 1: The table shows the best Manual ARMA and Auto-ARIMA models selected for Johnson & Johnson and Amazon stock data, with orders based on AIC values.

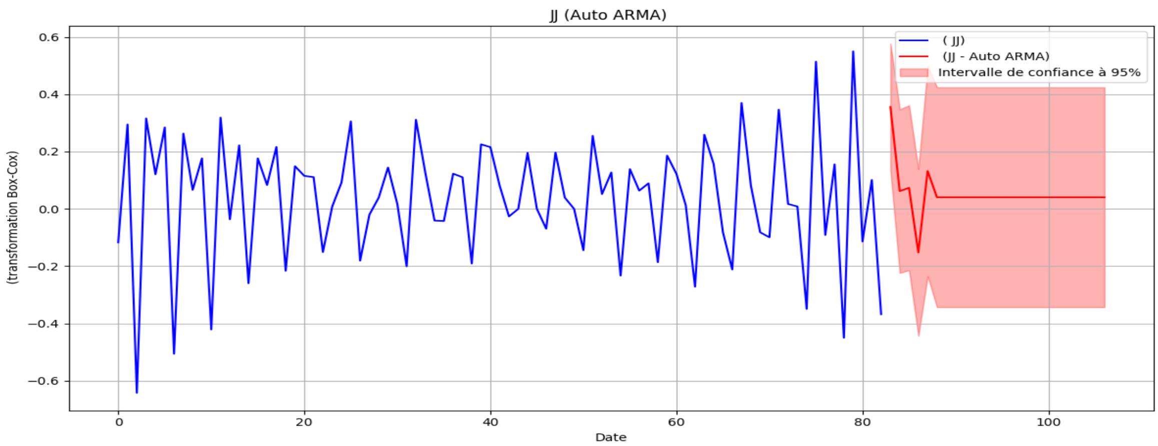
For **Johnson & Johnson Sales**, the **Manual ARMA model** with an order of **(3,0,2)** is chosen, which provides an AIC of **-144.35**, indicating a relatively good fit for the data. For **Amazon Stock Prices**, the **Manual ARMA model** with an order of **(0,0,0)** results in an AIC of **-10316.67**, showing a very simple model with a very low AIC

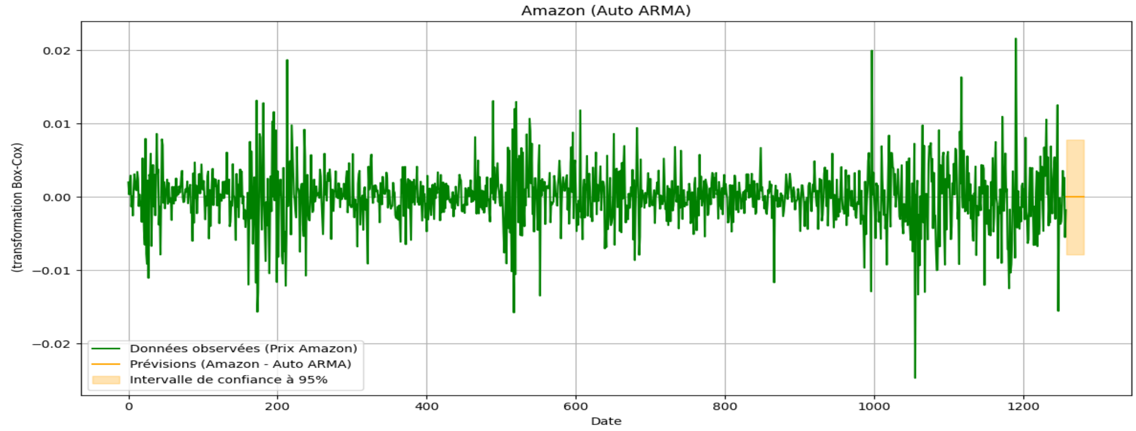




Graph plot5: (JJ&Amazon)forecast with confidence interval(Manual)

On the other hand, the **Auto ARIMA model**, with an order of **(0,0,5)**, has an AIC of **17.098**, which is higher, suggesting that the Auto ARIMA approach does not provide as efficient a model for this dataset. .while the **Auto ARIMA model**, also with an order of **(0,0,0)**, has an AIC of **1.644**, suggesting a slightly better fit for Amazon's stock price data.

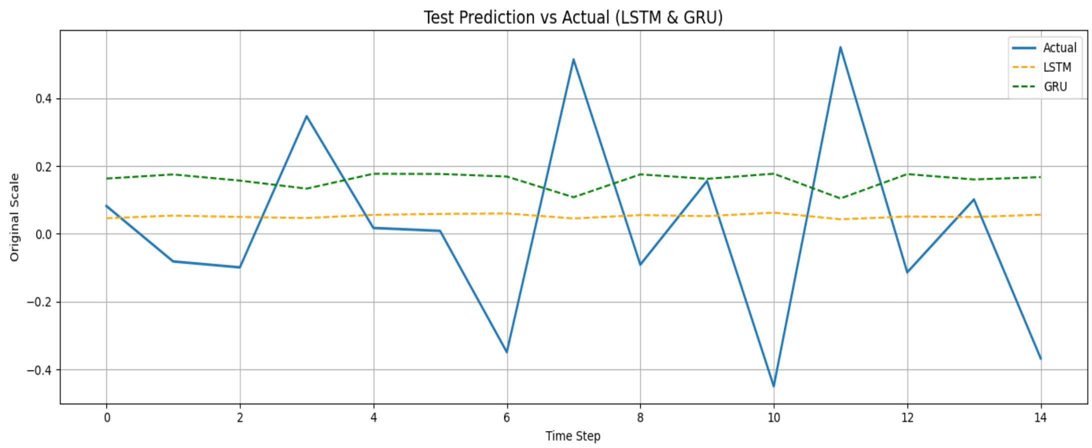


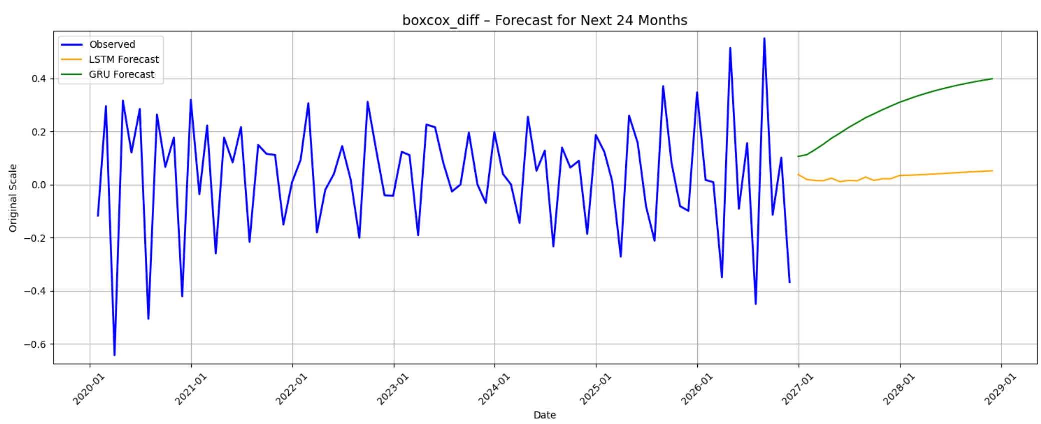


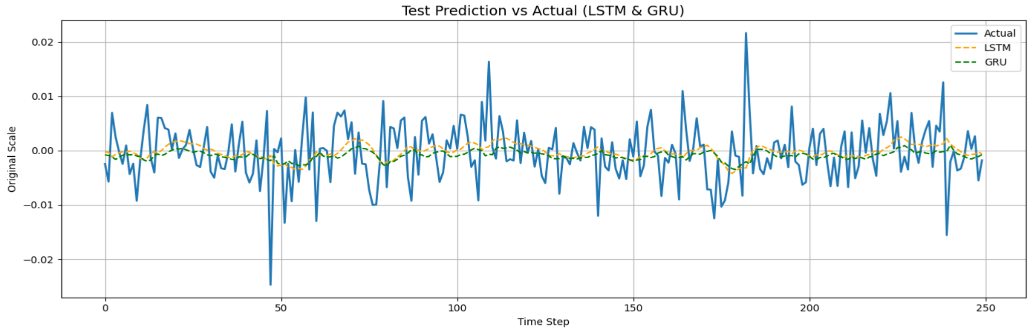
Graph plot6: (JJ&Amazon)forecast with confidence interval(Auto)

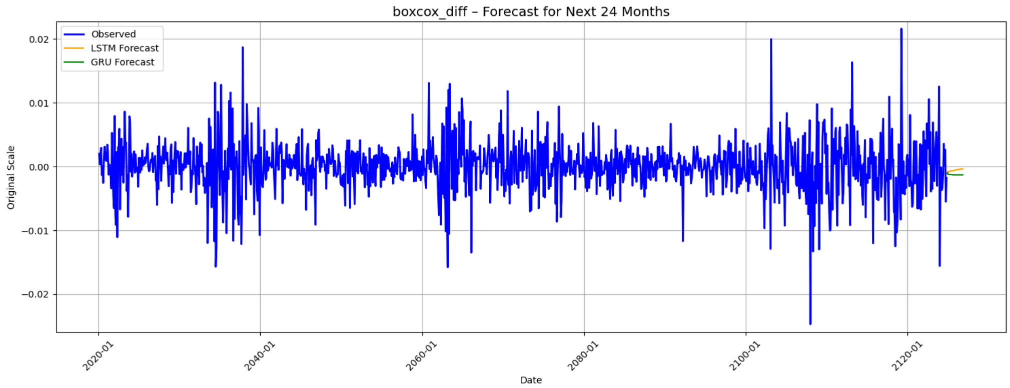
LSTM & GRU Time Series Forecasting

This script utilizes LSTM and GRU (models to forecast time series data, specifically for Johnson & Johnson sales and Amazon stock prices. The pipeline involves pre-processing the data by scaling and creating sequences of a fixed lookback period, followed by training the models on historical data. It uses MinMaxScaler for normalization and Early Stopping to prevent overfitting. After training, both models are evaluated using MAE**,** RMSE, and MAPE metrics, which help assess their accuracy. The script also incorporates recursive forecasting to predict future values, and it visualizes both test predictions versus actual values and the 24-month forecasts. Finally, it compares the performance of the LSTM and GRU models, providing insights into their forecasting capabilities for the given datasets.









Graphplot6: LSTM & GRU Time Series Forecasting

Final Discussion and Future Plan

The research successfully used LSTM and GRU models to forecast **Johnson & Johnson sales** and **Amazon stock prices**, demonstrating good predictive performance.

1. **Model Optimization**: Fine-tuning hyperparameters for better accuracy.
2. **Additional Features**: Incorporating external factors like economic indicators and market trends.
3. **Advanced Techniques**: Exploring Bidirectional LSTM, Attention Mechanisms, and hybrid models.
4. **Model Validation**: Using cross-validation for improved reliability.
5. **Real-Time Forecasting**: Implementing the models for continuous forecasting.

REFERENCE

1. ArunKumar, K.E., Kalaga, D.V., Sai Kumar, C.M., Kawaji, M. & Brenza, T.M., 2022. 'Title of the Article', *AEJ*, [online] Available at: https://doi.org/10.1016/j.aej.2022.01.011 [Accessed 11 April 2025]

2. Fu, R., Zhang, Z. & Li, L., 2017. 'Title of the Article', *IEEE Xplore*, [online] Available at: https://ieeexplore.ieee.org/document/7804912/authors#authors [Accessed 11 April 2025].
3. Zhao, L., Li, Z., Qu, L., Zhang, J. & Teng, B., 2023. 'Title of the Article', *Ocean Engineering*, [online] Available at: https://doi.org/10.1016/j.oceaneng.2023.114136 [Accessed 11 April 2025
4. Lingxiao, Z., Zhiyang, L., Leilei, Q., Junsheng, Z. and Bin, T., 2023. A deep learning-based model for offshore platform structure prediction. Ocean Engineering, [online] 286, p.114136. Available at: https://doi.org/10.1016/j.oceaneng.2023.114136

GITHUB LINK https://github.com/kamalibakthavatchalam/Time-series-modeling-/upload