Methodology

1 Exploration and Pre-processing

The dataset was imported from a CSV file and initially explored using functions like head() and str(). To enhance usability, The date column was formatted correctly as a Date type, while the wrong datatype column was converted to main datatype. Missing values were effectively managed through linear interpolation using the na.approx() function. For graphical representation use ggplot2. Perform eda.

2 Modeling and Diagnostic Techniques

To understand the data's autocorrelation structure, ACF and PACF plots were analyzed. A linear regression model was established as a baseline. An AR(1) model was subsequently tested, and an ARIMA(p,d,q) model was implemented to handle non-stationarity. Residuals were analyzed with the checkresiduals() function, and the Ljung-Box test checked for any remaining autocorrelation. To capture nonlinear dynamics, both TAR (Threshold Autoregressive) and STAR (Smooth Transition Autoregressive) models were applied. The Akaike Information Criterion (AIC) was used to compare model performanceand AIC and RMSE metrics were used to compare the performance of each model. Check model evaluation accuracy check of ML algorithm.

3 R Packages and Functions Used

The zoo package provided infrastructure for working with regular and irregular time series data. the ggplot2 package was employed for data visualization, particularly for creating complex plots. the forecast package provided methods and tools for forecasting time series data. the tseries package included tools for time series analysis. Use caret package for apply machine learning algorithm.