Journal of Intelligent Systems and Internet of Things
Vol. 17, No. 02, PP. 64-87, 2025
DOI: https://doi.org/10.54216/JISIoT.170206

Advanced Predictive Analysis of EGDI Time Series Using Hybrid ARIMA-LSTM and SARIMAX: A Comparative Study for Iraq and Tunisia

Ali Ahmed Ali[1,2,*], Atef Masmoudi[3]

[1]University of Sfax, National School of Electronics and Telecommunications of Sfax, Sfax, Tunisia

[2]Economic and Administration College, Al-Iraqia University, Baghdad, Iraq

[3]Laboratory of Electronics and Technology of Information, National Engineering School of Sfax, University of Sfax, Sfax, Tunisia

Emails: ali.ali@alIraqia.edu.iq; masatef@gmail.com

Abstract

This study presents a predictive modeling framework for forecasting the E-Government Development Index (EGDI) using two advanced time series approaches. Firstly, the Seasonal Auto Regressive Integrated Moving Average with Exogenous Variables (SARIMAX). Secondly, hybrid ARIMA-LSTM model. We focus on two case studies, Iraq and Tunisia, based on monthly EGDI data from the United Nations Survey Reports, spanning the years 2003 to 2024. Using several preprocessing steps such as handling missing data, testing for stationarity using the combined ADF and KPSS tests, and determining the optimal ARIMA parameters through ACF and PACF analysis and implementing autoarima. The model was built and trained using 80% of the data, while 20% was retained for testing. The independence of the residuals verified using the Ljung-Box test. Four types of visualization and error analysis were applied using ACF/PACF for residuals, error plots as prediction error plot, error distribution plot (histogram + KDE) and decomposition analysis to visually assess model fit. Evaluation was conducts using multiple error metrics, including RMSE, MAE, MAPE, MHE, AIC, BIC and MAPA. After building the four models, we ensured that the results and reconstructions were evaluated using the 12 tests we mentioned, and that they were based on the best results and were consensus acceptable. ARIMAX model demonstrated superior performance, achieving an average absolute percentage Accuracy (MAPA) of 98.35% for Iraq and 97.93% for Tunisia. In comparison, the hybrid ARIMA-LSTM model, which combines linear ARIMA outputs with nonlinear corrections from an LSTM neural network, demonstrated competitive predictive ability with a MAPA of 95.68% for Iraq and 96.14% for Tunisia. SARIMAX showed slightly outperformed the hybrid model in overall accuracy. On other hand, ARIMA-LSTM model demonstrated robustness in capturing complex nonlinear dynamics particularly in the more structurally diverse Tunisian dataset. These results confirm the potential of both models as effective tools for predicting EGDIs and support their application in digital governance planning and policymaking. We designed and we recommend adopting our "12 -Test Approach" for evaluation framework as a standard methodology in future studies addressing analysis and forecasting, and its suitability for different types of time series models. This approach provides comprehensiveness, accuracy, and flexibility in evaluation, regardless of model type or application area.

Received: January 31, 2025 Revised: March 28, 2025 Accepted: June 04, 2025

Keywords: SARIMAX; ARIMA; LSTM; e-government; EGDI; predictive modeling, Iraq, Tunisia; Time-Series

1. Introduction

E-Government means the obligation of national authorities providing number of online services relying on new communication technologies [1]. The United Nations measured the readiness and capacity of E-Government Development index (EGDI) relying on the information and communications [2]. EGDI is a nascent field that urgently requires modelling then predicting. Forecasting the EGDI requires historical data that are relay to be readily available, as the first UN e-Government Survey was stared in 2003 [1], while the latest edition published in 2024 [3]. Thus, EGDI requires reliable forecasting tools for strategic planning. In the Middle East, Iraq and Tunisia exhibited distinct patterns in the evolution of EGDI motivating comparative forecasting study. Data from Iraq and Tunisia recollected and processed. Two decades of time series data spanning 53 months used to model and forecast the EGDI. In this paper, it is aimed to build a robust forecasting model by using hybrid techniques SARIMAX and ARIMA-LSTM, and evaluating effectiveness.

1.1 Hybrid Arima-Lstm Model

Deep learning frameworks and statistical time series models are used to create hybrid models that can capture both linear and nonlinear dynamics in complex datasets. The hybrid ARIMA-LSTM model combines autoregressive integrated moving averages (ARIMA) with long-short-term memory (LSTM) neural networks. This hybrid strategy combines the benefits of both approaches. ARIMA models the time series' linear structure, while LSTM captures its nonlinear interactions.[4], [5]. The hybrid forecast at time t, denoted as $\hat{\mathcal{Y}}_t$, is given by: $\hat{\mathcal{Y}}_t = \hat{\mathcal{Y}}_t{}^{ARIMA} + \hat{e}_t{}^{LSTM}$

Where:$\hat{\mathcal{Y}}_t$: hybrid forecast, $\hat{\mathcal{Y}}_t{}^{Arima}$: ARIMA forecast and $\hat{e}_t{}^{LSTM}$: LSTM forecast of the residual

This two-stage forecasting framework is known as the hybrid ARIMA-LSTM model. To account for the linear nature of the time series, an ARIMA model is first fitted to the data. An LSTM neural network is then used to simulate the residuals (errors) from the ARIMA model, which are believed to contain nonlinear information. The predicted residuals from the LSTM and the ARIMA predictions are combined to obtain the final forecast [6] . To achieve the better prediction results, we constructed a hybrid model that combines the advantages of ARIMA and LSTM. The autopilot data, being time series data, can be assumed to consist of both linear and nonlinear components, represented as follows: $X_t = L_t + N_t + \varepsilon_t$

where $L_t$represents the linearity components of the data at time t which processed with ARIMA, $N_t$represents the nonlinearity components which processed with LSTM, and εt represents the error term [7]. The ARIMA model captures the linear component of a time series using three parameters: p: the rank of the autoregressive part, d: the degree of divergence, and q: the rank of the moving average part [4]. An LSTM model uses residual sequences to train the LSTM model. The residual series $e_t = y_t + y_t{}^{Arima} + \varepsilon_t$ is used to train an LSTM model. The LSTM captures non-linear dependencies using gated memory cells and is trained over time-lagged sequences of residuals [8]. The forget, input, output and cell state update gates in LSTM equations control the flow of temporal information. Hybrid ARIMA-LSTM model applications are domains in which time series data has varying and complex nonlinear dynamics over time and include linear trend information. Moreover, applications exist in realms of banking, energy, health, and economy [5]. The hybrid model also has several advantages, such as: (1) good performance on complex datasets, (2) flexible in processing multivariate time series and, (3) ability to detect both linear and nonlinear patterns. The challenges lie in the need for tuning, sensitivity to scale and season, and higher computational costs. The ARIMA-LSTM hybrid model draws best from both worlds by combining ARIMA's statistical accuracy with the deep learning capability brought by LSTMs to develop an effective and powerful time series-forecasting tool. Its capacity to imitate a wide range of data patterns makes it perfect for many real-world applications. Scientific procedures are needed for the ARIMA model, such as inductive testing, the kpss test, or the augmented Dickey-Fuller (ADF) test to confirm stability [9]. Finding the optimal parameters (p, d, and q) [10] by used the partial autocorrelation factor (PACF) and autocorrelation factor (ACF). Applying maximum likelihood estimation (MLE) techniques to estimate the model [11]. Evaluation using quantitative model indicators such as RMSE, AIC, and BIC measures [12]. Transforming ARIMA equations and applying them practically using Python (statsmodels package) and other programming languages [13], [14]. In addition to the direct benefits and uses of ARIMA models for creating predictive models for financial markets and forecasting the price of the dollar, currencies, and other corporate stocks [15], [16]. forecasting, economic forecasting, GDP forecasting, and most aspects of the national economy [17], [18]. Climate forecasting, weather forecasting, drought, wind speed, and rainfall [19], [20]; forecasting fuel consumption and demand for gas and electricity [21], [22]. The success of the ARIMA model in various fields has necessitated the modification, development, and construction of other models based on it. Seasonal and multi-decomposition models (SARIMA), SARIMAX, hybrid ARIMA, and artificial neural network (ANN) models have been developed [20], [22]. Hybrid ARIMA-ANN models have been able to capture patterns and improve forecasting accuracy [23]. We can say that the most relied upon model that has contributed to the development and success of time series forecasting is the ARIMA model.

1.2 Sarimax Model

The Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) is basically an extension of the ARIMA model to include seasonal effects as well as external variables. This model is applicable where the time series data show seasonal patterns and the external factors influence them [24]. The components that characterize the SARIMAX model are based on [25].Non-seasonal parameters (p, d, q) parameter (p) represents order of the autoregressive part, (d) represents degree of first differencing involved and (q) represents order of the moving average part. Seasonal parameters (P, D, Q, s) order of the seasonal autoregressive part represent by (P), degree of seasonal differencing represent by (D), order of the seasonal moving average part represent by (Q) and (s) represents length of the seasonal cycle. Exogenous variables inclusion of external variables that may influence the target time series.

The general form of the SARIMAX model can be expressed as:

$$\Phi_P(B^s)\ \varphi_P\,(B)(1-B)^d(1-B^s)^D Y_t = \Theta\_Q(B^s)\ \theta\_q(B)\ \varepsilon_t + \beta X_t$$

Where:

- Y_t is the observed value of the target series at time t.
- B is the backshift operator.
- $\varphi_P\,(B)$ and $\theta_q(B)$represent the non-seasonal AR and MA polynomials.
- $\Phi_P(B^s)$ and $\Theta\_Q(B^s)$ represent the seasonal AR and MA components.
- ε_t denotes the white noise error term.
- X_t is a vector of exogenous variables.
- β are the coefficients associated with the exogenous inputs.

SARIMAX models are widely used in various fields, including [26]

- Economics: Forecasting indicators like GDP, inflation rates, and unemployment, considering external economic factors.
- Energy Sector: Predicting electricity consumption and generation by incorporating weather conditions as exogenous variables.
- Retail Industry: Forecasting sales by including promotions, holidays, and other external events.
- Public Health: Modeling disease outbreaks by considering environmental factors.

2. Methodology

The e-government indicators EGDI, HCI, OSI and TII data are collected from two countries of Iraq and Tunisia through the UN survey versions for the years 2003, 2005, 2008, 2010, 2012, 2014, 2016, 2018, 2020, 2022 and 2024 [1]–[3], [27]–[34]. Finally, two individual datasets were created for each of the countries under study. The main purpose is to reveal the performance expectations of e-government indicators according to times periods, each dataset contained 253 instants. This research aims to build an SARIMAX and ARIMA-LSTM models to predict the e-government index (EGDI) data for Iraq and the Tunisia using Py-thon language. Time series modelling methodology is one of the most important tools in analysing e-government data The procedure commences with the installation of fundamental libraries, including pandas, statsmodels, and pmdarima. This study employs a hybrid methodology utilizing ARIMA and long- and short-term trend-based neural networks (LSTM) to enhance the time series prediction of the E-Government Development Index (EGDI) for Iraq and Tunisia. Following stationarity assessments via the ADF and KPSS tests, along with requisite differentiation to stabilize the data. Using both tests ADF and KPSS together provides a more robust view of the stationarity of the series. The hypred ARIMA model was chosen with best parameters (3, 1, 2) for IRAQ and TUNISIA. While SARIMAX model for Iraq adopted ordering (1, 1, 0), (1,1,0,12) and for Tunisian (3, 1,2), (1,1,1,12). The model was subsequently trained on the EGDI data following the division of the dataset into a training set (80%) and a testing set (20%). After extracting the ARIMA model predictions and calculating the residuals, these residuals were divided into fixed-length subseries of 20 time points to train the LSTM model, which consisted of 256 neurons using a dense layer and a dropout layer with a value of 0.1. The model was trained for 200 training epochs. This methodology was applied to EGDI data for both Iraq and Tunisia, helping to improve forecast accuracy by leveraging ARIMA's linear prediction capabilities and LSTM's ability to handle nonlinear variations in time series, enhancing the accuracy of the results and the robustness of the proposed hybrid model. Both models use 10 libraries, with two differences: the SARIMAX model uses the pmdarima library. The hybrid ARIMA-LSTM model uses the TensorFlow library (specifically, the Keras library with LSTM layers). Besides, configuring the necessary work environment, then the dataset is loaded from the Excel file and the required columns are checked (for example "year" and "EGDI"). The "year" column is then converted to date time format and used as an index for the data with a monthly frequency set to ensure the regularity of the time series.

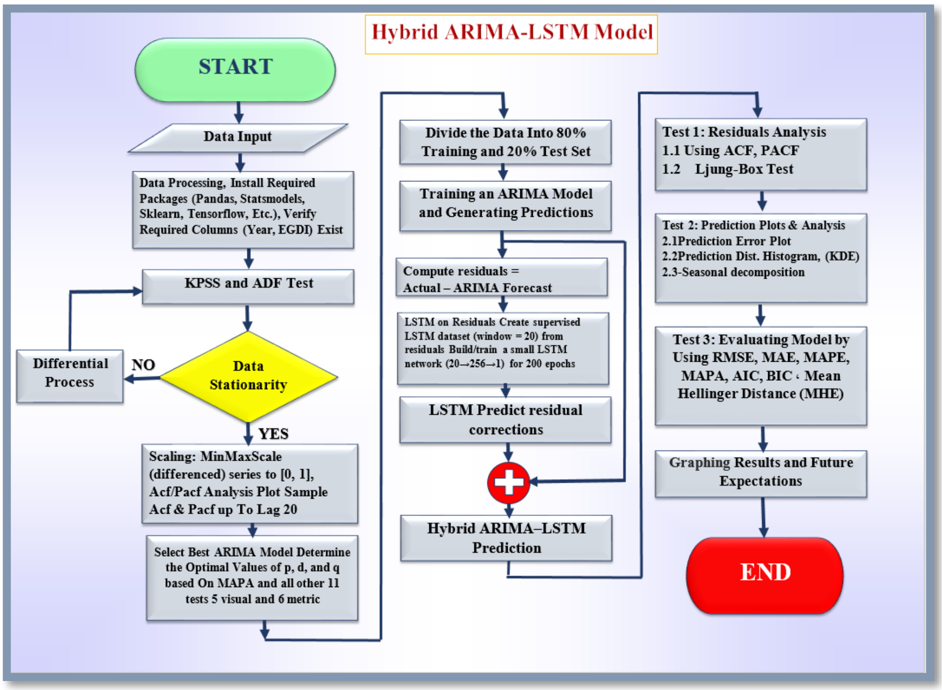

Figure 1. Methodology Flow Chart for Hybrid ARIMA-LSTM Model

The methodology used in this study on the EGDI data for Iraq and Tunisia relies on a hybrid model that combines ARIMA and long- and short-term trend-based neural networks (LSTM) to improve the performance of time series prediction of the E-Government Development Index (EGDI). The ARIMA model was selected with values (3, 1, 2) after conducting stationarity tests using the ADF and KPSS tests and applying the necessary differentiation to stabilize the data. Then, the model was trained on EGDI data after splitting it into a training set (80%) and a testing set (20%). After extracting the predictions from the ARIMA model and calculating its residuals, those residuals were divided into fixed-length subseries of 20 time points for training the LSTM model. This model comprised 256 neurons in a dense layer accompanied by a dropout layer whose value is 0.1, trained for 200 epochs. It applies this methodology to EGDI data for Iraq and Tunisia; hence, it improves forecasting accuracy by using ARIMA's strength in linear prediction along with LSTM's ability to learn non-linear dynamic changes within time series making the results more accurate as well as having a more robust hybrid model.

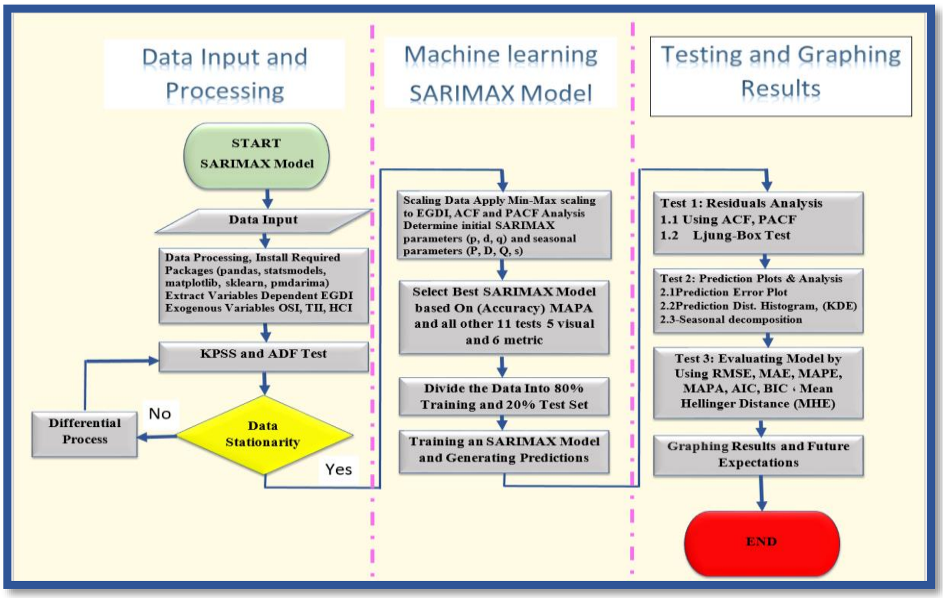

Figure 2. Methodology Flow Chart for SARIMAX Model

The flow of work in the construction of the SARIMAX model initiates with data collection for the years 2003-2024. In the first step, data is cleaned dealing with missing values by interpolating mean etc., and removing outliers that could affect negatively on the accuracy of a model. A SARIMAX model was built to predict the E-Government Development Index (EGDI) via a methodology that mixes indigenousness of time series with seasonal exogenous impacts. In this model, stationarity of the series was checked using ADF and KPSS tests; differentiation was applied to make stationary and to remove general trends. Data has been normalized using Min-Max methodology to make stability of the model better and bring quality predictions. The SARIMAX model for Iraq adopted a non-seasonal ordering (1, 1, 0), which means using a first-order autoregressive component with a single differentiation to remove linear trends, without using a moving average component. In addition, a seasonal ordering (1, 1, 0, 12) was used, which indicates annual (12-month) seasonality with a first-order seasonal autoregressive component and a first-order seasonal differentiation to remove seasonal effects. In contrast, the Sarimax model was applied to Tunisia with a non-seasonal (3, 1, 2) order using three autoregressive coefficients, one differential and two moving average coefficients with a seasonal order (1, 1, 1, 12), including seasonal autoregressive, seasonal differential, and seasonal moving average coefficient, in addition to a 12-month seasonal cycle. The data were split into a training, test set to evaluate performance, and the residuals were analyzed using the Ljung-Box test and ACF, and PACF plots to confirm their independence and model effectiveness. The performance of the models was evaluated using various metrics such as RMSE, MAE, MAPE, AIC, BIC, correlation coefficient, and MAPA, confirming the efficiency of the SARIMAX model and its high ability to predict EGDI values in both countries Iraq and Tunisia.

3. Data Preparation

3.1 Data Stationarity Test Using (ADF) and KPSS Test

ADF- Augmented Dickey-Fuller statistical test assesses the existence of a unit root in a time series to see if the series is stable around a constant mean, indicating the presence of a unit root. It is articulated as follows: $\Delta y_t = \alpha + \gamma y_{t-1} + \varepsilon_t$ [10].

While KPSS- Kwiatkowski–Phillips–Schmidt–Shin. the null hypothesis is that the series is stationary. In the ADF test, the null hypothesis is that the series is nonstationary (contains a unit root) [8]. We used KPSS and ADF together because using only one test (ADF or KPSS) can sometimes be inconclusive or misleading, because each test tests a different hypothesis. However, using both tests together provide a more robust view of the stationarity of the series.

3.1.1 Iraq ADF and KPSS Test Results

Table 1: Iraq Augmented Dickey-Fuller (ADF) and KPSS Test

Augmented Dickey-Fuller (ADF) and KPSSTest Results for IRAQ EGDI

Test Type	Statistic	p-value	Lags Used	Observations	Critical Value 1%	Critical Value 5%	Critical Value 10%	Stationarity
ADF (diff 0)	-1.6805	0.4412	1	251	-3.4567	-2.8731	-2.5729	Not stationary
ADF (diff 1)	-24.899	0	0	251	-3.4567	-2.8731	-2.5729	Stationary
KPSS	0.2999	0.1	8	250	0.739	0.463	0.347	Stationary

The table displays the results of stationarity tests for the EGDI series in Iraq, utilizing the Augmented Dickey-Fuller (ADF) and KPSS methods. The ADF test results at the (diff 0) level show that the statistical value (-1.6805) surpasses the critical values at all levels, accompanied by a p-value of 0.4412. This indicates that the hypothesis of a unit root cannot be rejected, suggesting that the series is non-stationary. After first-order differentiation, the statistical value of -24.899 was significantly lower than the critical values, with a p-value of 0.000, indicating the stability of the series following differentiation. The KPSS test yields a statistical value of 0.2999, which is below the critical values at all levels, suggesting acceptance of the null hypothesis of stationarity.

3.1.2 Tunisia ADF and KPSS Test Results

Table 2: Tunisia Augmented Dickey-Fuller (ADF) and KPSS Test

Augmented Dickey-Fuller (ADF)and KPSS Test Results for Tunisia EGDI								
Test	Statistic	p-value	Lags Used	Observations	Critical Value 1%	Critical Value 5%	Critical Value 10%	Stability Status
ADF (diff 0)	-0.8853	0.7928	1	251	-3.457	-2.873	-2.573	Not stationary
ADF (diff 1)	-2.827	0.0545	0	251	-3.457	-2.873	-2.573	Borderline
ADF (diff 2)	-15.752	0	0	250	-3.457	-2.873	-2.573	Stationary
KPSS	0.051	0.1	0	250	0.739	0.463	0.347	Stationary

The EGDI series in Tunisia underwent stationarity tests using ADF and KPSS tests. The ADF (diff 0) value is -0.8853 with a p-value of 0.7928, indicating no stationary at its original level. The ADF (diff 1) value is -2.827 at the 10% level (-2.573) and above the critical value at the 5% level (-2.873), with a p-value of 0.0545, borderline, and stationary after the second differentiation. The KPSS value is 0.051, confirming the null hypothesis of stationarity, supporting the series's stationarity after processing. The EGDI series for Tunisia is non-stationary at its initial level, showing an inconclusive stationarity after the first differentiation, but it becomes clearly stable after the second differentiation. Therefore, a more appropriate ARIMA or SARIMAX model requires the use of a second-order difference (d = 2).

3.2 ACF and PACF Test

Autocorrelation and partial correlation functions offer visual tools for understanding time series correlation structure, diagnosing dynamic properties, and determining autoregressive components. ACF calculates correlation coefficients and PACF evaluates direct relationship between series and lagged values. [4]

3.2.1 Iraq (ACF) Test and (PACF) Test

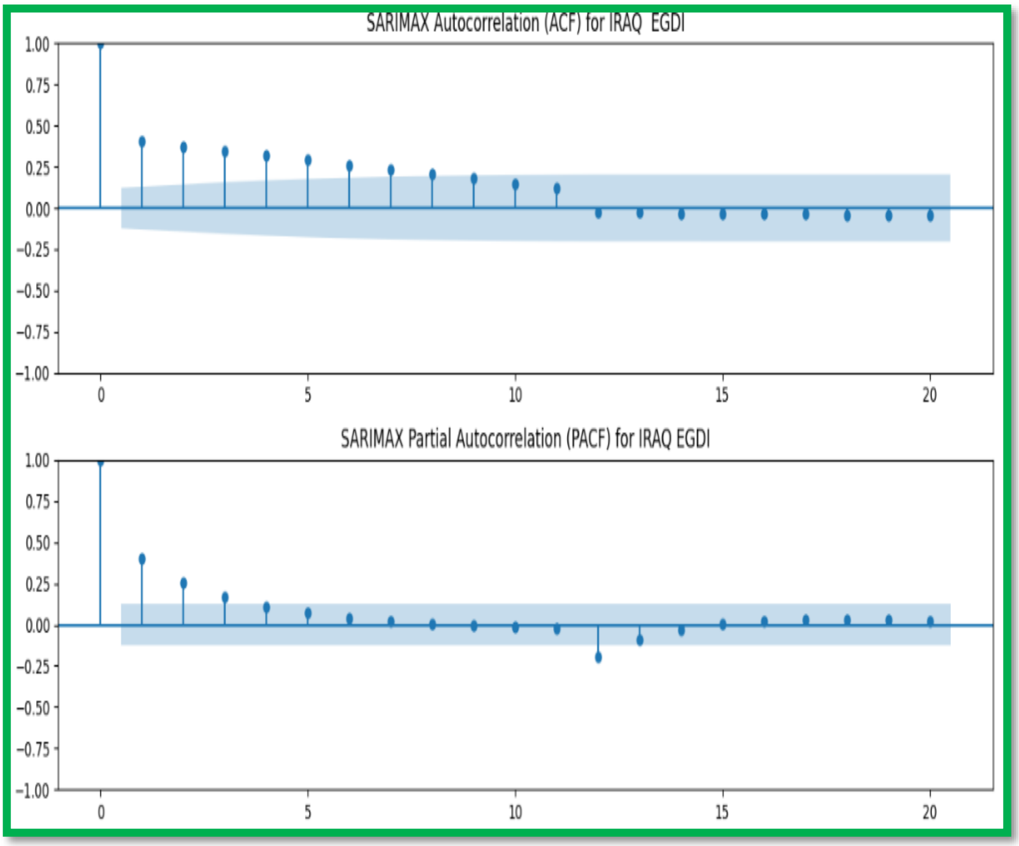

Figure 3. Iraq ACF and PACF Test

The ACF and PACF plots for Iraq's EGDI indicate that the original series is non-stationary, as evidenced by the gradual decay in the ACF. The ACF exhibits notable positive autocorrelations in the initial lags, which progressively diminish, characteristic of an integrated process. The PACF shows a sharp cut off after lag 1, implying the presence of a strong autoregressive component of order 1. Together, these patterns support using a SARIMAX and ARIMA-LSTM model with an AR term of order 1 and a first-order difference, indicating a suitable structure like ARIMA (1,1, q) or SARIMAX (p=1, d=1, q).

3.2.2 Tunisia (ACF) Test and (PACF) Test

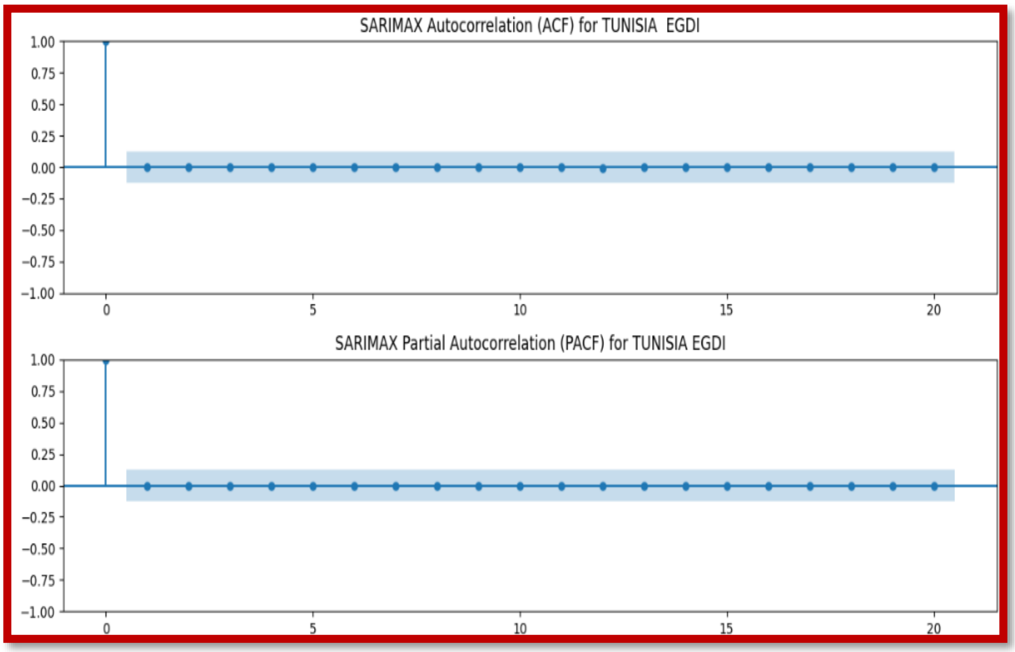

Figure 4. Tunisia Autocorrelation Function (ACF) Test

The ACF and PACF plots for Tunisia's EGDI show no significant spikes beyond lag 0, indicating that the series is white noise after differencing. This evidence suggests that the SARIMAX and ARIMA-LSTM model used has effectively removed autocorrelation, and the residuals are stationary and uncorrelated. Therefore, the model is likely well specified.

4. Training and Validation

Both models SARIMA and ARIMA-LSTM splits the data series into 80% training data and 20% for testing and validation data Sarimax model use Exogenous variables (exog) HCI, OSI and TIII are used at each stage to ensure the accuracy and reliability of the forecasts. Both models we used for models updated gradually by adding a new point after each prediction. This type of training was based on a "rolling forecast" technique. After obtaining the differential and matched series predictions, the initial values were restored by cumulative summing and demoralization using the MinMaxScaler arithmetic inverse.

5. Testing and Graphing Results

We relied on 12 main tests in our study to verify our work divided it in three main test group as follow. Residuals Analysis (ACF), PACF, and Ljung-Box. Second, Prediction Visualization and Distribution contain a Prediction Error Plot, a Distribution Histogram, (KDE) analysis of errors, and Time Series Decomposition (Additive) analysis. Third, Quantitative Evaluation (QE), MAE, MAPE, RMSE, AIC, BIC, MHE, and Accuracy (MAPA).

5.1 Test 1 Residuals Analysis by ACF, PACF and LJUNG-BOX TEST

Test 1 focuses on statistical analysis of residuals such as the Ljung-Box test and ACF/PACF analysis and has two objectives testing the independence of errors and the absence of temporal correlation.

5.1.1 ACF and PACF Test for SARIMAX MODEL

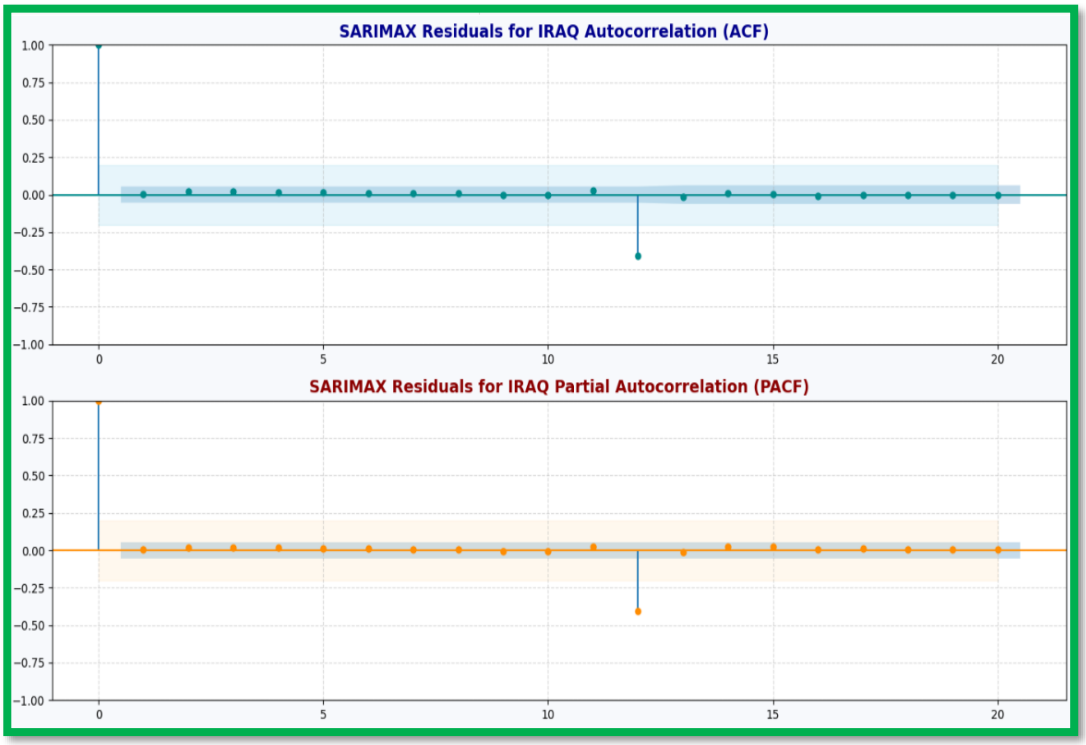

Figure 5. IRAQ SARIMAX Residuals (ACF) and (PACF) Test

The ACF plot for Iraq shows that all values except lag = 0 fall within the shaded band (95% confidence interval), indicating that there is no significant autocorrelation in the residuals. This evidence indicates that the SARIMAX model captures the time structure of the series well and leaves no unexplained temporal pattern. The PACF plot reveals that there is no correlation between the residuals at each lag, eliminating the influence of intermediate lags. Most values appear within the confidence interval, reinforcing the idea that the residuals do not exhibit direct time dependence. The PACF plot confirms the ACF results and indicates that the residuals do not suffer from lag dependence. This supports the model structure from a time dynamics perspective.

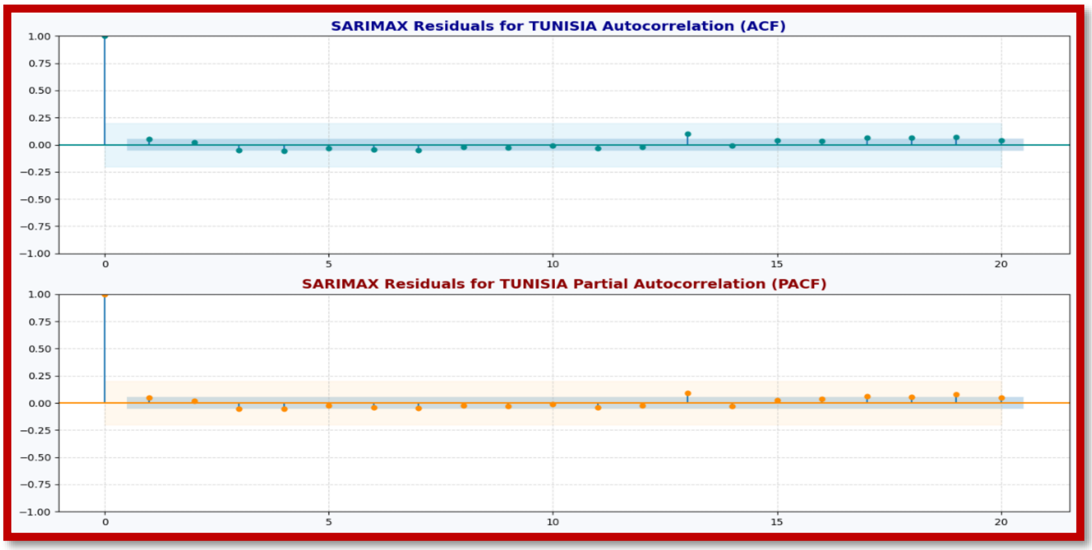

Figure 6. Tunisia SARIMAX (ACF) and (PACF) Test

In Fig 6. above shows Tunisia ACF plot of the residuals (top panel) displays values that fall well within the 95% confidence bounds across all lags, except for lag 0 which is expected to spike due to self-correlation. The absences of significant spikes indicate that the residuals are uncorrelated and resemble white noise, implying that no systematic information is left unmodelled. Likewise, the PACF plot (bottom panel) confirms the same conclusion. The lags do not show clear partial autocorrelations beyond the 95% bounds. The model has adequately captured the underlying structure of the time series. The lack of serial correlation in both plots confirms the validity of the SARIMAX specification, and that the residuals are statistically independent and identically distributed.

5.1.2 ACF and PACF for Hybrid ARIMA-LSTM

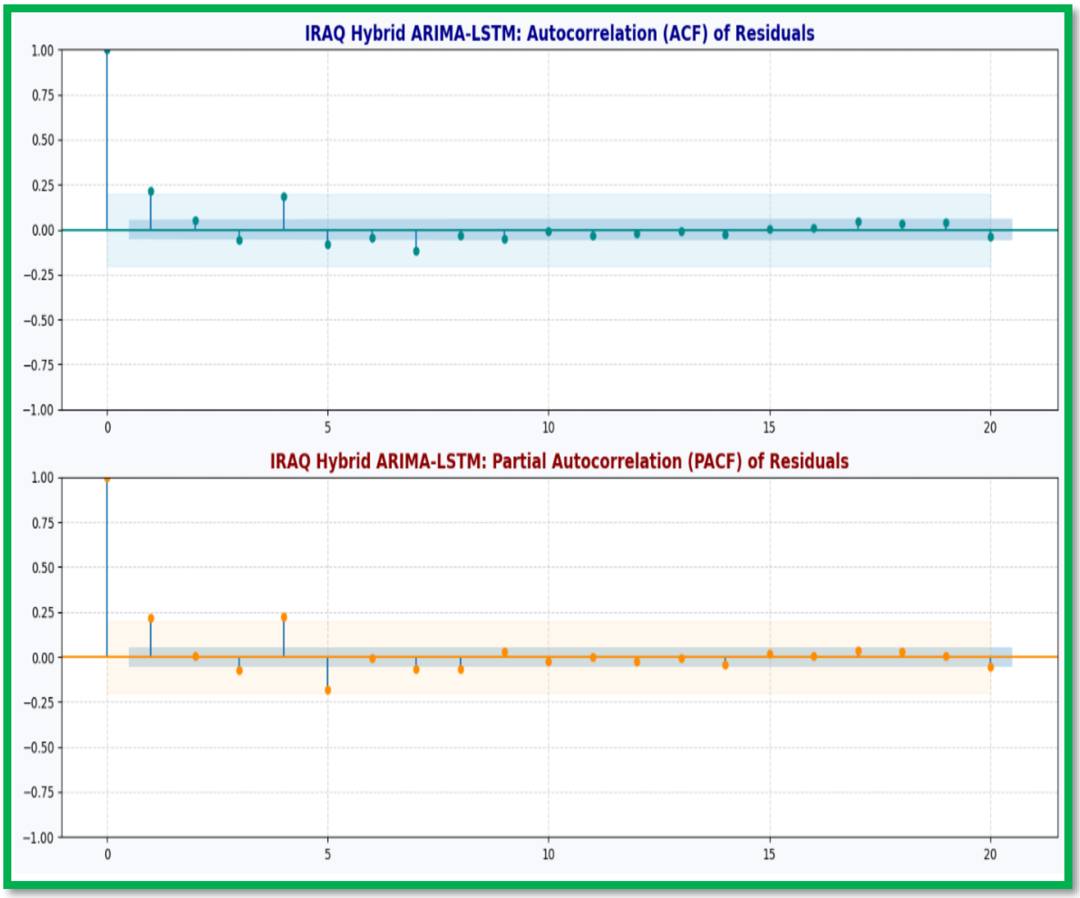

Figure 7. Iraq Hybrid ARIMA-LSTM ACF and PACF Test

IRAQ Hybrid ARIMA-LSTM Residuals plot for Iraq's (ACF) as shown in the top figure most correlations fall within the confidence limits (the cyan shaded area) reflects the absence of autocorrelation for most lags. The model was able to adequately extract the temporal patterns present in the data and left only a small effect in the residuals. Only a slight lag is visible at lag = 1, which is statistically acceptable as long as it is outside the significance limits. Residuals plot for Iraq's (PACF) the bottom figure represents partial autocorrelation and shows the same statistical behaviour, with the majority of values falling within the confidence limits, and there are no strong indications of a linear pattern structure not captured by the model.

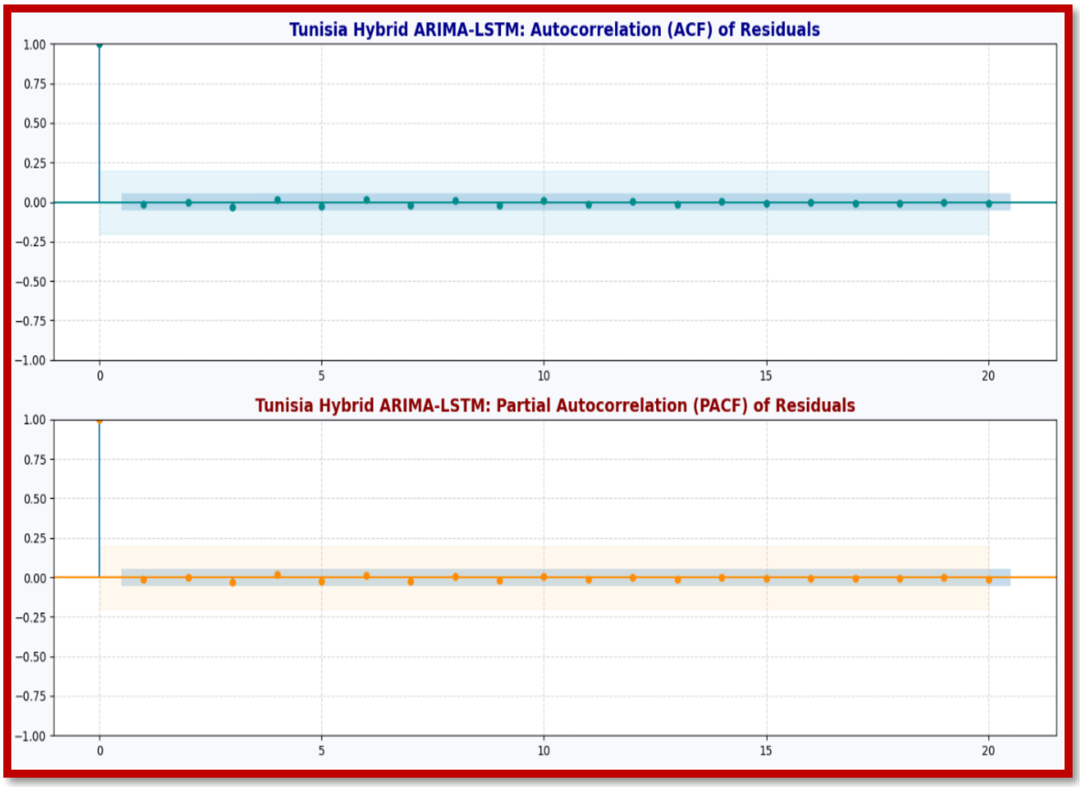

Figure 81. Tunisia Hybrid ARIMA-LSTM ACF and PACF Test

The figure 8 shows the results of (ACF) and (PACF) analyses of the residuals of the Tunisia ARIMA-LSTM hybrid-forecasting model. All values fall within the 95% confidence intervals, according to the ACF and PACF plots. There are no statistically significant temporal correlations between the residuals over different times. This stochastic behaviour of the residuals supports the hypothesis that the model successfully extracted the underlying structural information in the original time series, as no unrepresented temporal pattern remained. According to the statistical criteria used to evaluate the effectiveness of models, the apparent pattern of the residuals indicates the model's suitability for prediction and the absence of systematic variations that were not captured, which enhances the reliability and accuracy of the model in future applications. RESIDUALS ANALYSIS BY LJUNG-BOX TEST

5.1.3 Ljung-Box test for SARIMAX Model

Table 3: Iraq SARIMAX Ljung-Box Test

SARIMAX: IRAQ Ljung-Box Test Results					
Lag	**p-value**	**Result**	**Lag**	**p-value**	**Result**
1	0.9679	Independent Residuals	**11**	1	Independent Residuals
2	0.988	Independent Residuals	**12**	0.4842	Independent Residuals
3	0.9973	Independent Residuals	**13**	0.5647	Independent Residuals
4	0.9995	Independent Residuals	**14**	0.642	Independent Residuals
5	0.9999	Independent Residuals	**15**	0.7123	Independent Residuals
6	1	Independent Residuals	**16**	0.7735	Independent Residuals
7	1	Independent Residuals	**17**	0.8258	Independent Residuals
8	1	Independent Residuals	**18**	0.8689	Independent Residuals
9	1	Independent Residuals	**19**	0.9034	Independent Residuals
10	1	Independent Residuals	**20**	0.9303	Independent Residuals

Table above IRAQ SARIMAX Ljung-Box test show p-values exceed the conventional significance threshold (0.05), indicating that the null hypothesis of independence cannot be rejected. Thus, the residuals exhibit no significant autocorrelation and are considered white noise, supporting the adequacy of the model.

Table 4: TUNISIA SARIMAX Ljung-Box TEST

SARIMAX: TUNISIA Ljung-Box Test Results					
Lag	**p-value**	**Result**	**Lag**	**p-value**	**Result**
1	0.7068	Independent Residuals	**11**	0.9999	Independent Residuals
2	0.921	Independent Residuals	**12**	1	Independent Residuals
3	0.9544	Independent Residuals	**13**	0.9999	Independent Residuals
4	0.9698	Independent Residuals	**14**	1	Independent Residuals
5	0.9883	Independent Residuals	**15**	1	Independent Residuals
6	0.9944	Independent Residuals	**16**	1	Independent Residuals
7	0.9969	Independent Residuals	**17**	1	Independent Residuals
8	0.9989	Independent Residuals	**18**	1	Independent Residuals
9	0.9996	Independent Residuals	**19**	1	Independent Residuals
10	0.9999	Independent Residuals	**20**	1	Independent Residuals

Table3 above Tunisia SARIMAX Ljung-Box the p-values for all lags (1 through 20) indicate that null hypothesis of independence cannot be rejected. Thus, the residuals exhibit no significant autocorrelation and are considered white noise, supporting the adequacy of the model.

5.1.4 Ljung-Box Test for Hybrid ARIMA-LSTM Model

Table 5: Iraq Hybrid ARIMA-LSTM Ljung-Box TEST

IRAQ Hybrid ARIMA-LSTM Ljung-Box Test Results					
Lag	**p-value**	**Result**	**Lag**	**p-value**	**Result**
Lag 1	0.1122	Independent Residuals	Lag 11	0.84	Independent Residuals
Lag 2	0.2623	Independent Residuals	Lag 12	0.8881	Independent Residuals
Lag 3	0.4121	Independent Residuals	Lag 13	0.9252	Independent Residuals
Lag 4	0.3048	Independent Residuals	Lag 14	0.9501	Independent Residuals
Lag 5	0.3926	Independent Residuals	Lag 15	0.9685	Independent Residuals
Lag 6	0.5048	Independent Residuals	Lag 16	0.9806	Independent Residuals
Lag 7	0.5206	Independent Residuals	Lag 17	0.9865	Independent Residuals
Lag 8	0.6207	Independent Residuals	Lag 18	0.9913	Independent Residuals
Lag 9	0.6993	Independent Residuals	Lag 19	0.9943	Independent Residuals
Lag 10	0.7803	Independent Residuals	Lag 20	0.9963	Independent Residuals

The table above fore Hybrid ARIMA-LSTM model clearly summarizes the for residual independence across 20 lags for the. All p-values are above 0.05 supporting the null hypothesis of white noise residuals. This confirms that the residuals are not auto correlated and behave as a random sequence indicating that the model has effectively captured the time-dependent structure in the data.

Table 6: Tunisia Hybrid ARIMA-LSTM Ljung-Box TEST

TUNISIA Hybrid ARIMA-LSTM Ljung-Box Test Results					
Lag	p-value	Result	Lag	p-value	Result
Lag 1	0.9197	Independent Residuals	Lag 11	1	Independent Residuals
Lag 2	0.9948	Independent Residuals	Lag 12	1	Independent Residuals
Lag 3	0.9959	Independent Residuals	Lag 13	1	Independent Residuals
Lag 4	0.9992	Independent Residuals	Lag 14	1	Independent Residuals
Lag 5	0.9997	Independent Residuals	Lag 15	1	Independent Residuals
Lag 6	1	Independent Residuals	Lag 16	1	Independent Residuals
Lag 7	1	Independent Residuals	Lag 17	1	Independent Residuals
Lag 8	1	Independent Residuals	Lag 18	1	Independent Residuals
Lag 9	1	Independent Residuals	Lag 19	1	Independent Residuals
Lag 10	1	Independent Residuals	Lag 20	1	Independent Residuals

The column table of the Ljung-Box test results for the Hybrid ARIMA-LSTM model for Tunisia reflects the robust statistical performance of the model's residuals. The very high p-values, ranging from 0.9197 to 1.0000 across all lags (lags 1 to 20), show that the model's residuals do not contain statistically significant temporal autocorrelation, indicating that these residuals are independent.

5.2 TEST 2: Prediction Plots & Analysis

5.2.1 A prediction error plot Test

5.2.1.1 A prediction Error Plot Test for SARIMAX Model

An analytical tool used to evaluate the performance of a forecasting model by displaying the differences between actual and predicted values over time. Errors are represented on the vertical axis (Actual - Predicted) while time intervals are represented on the horizontal axis (Time Steps). Beside, Prediction Error Distribution Analysis using a Histogram and Kernel Density Estimation (KDE) used to evaluate how well a forecasting model aligns with actual values by analyzing mean error, standard deviation of Errors, probability distribution and (KDE) - Kernel Density Estimation.

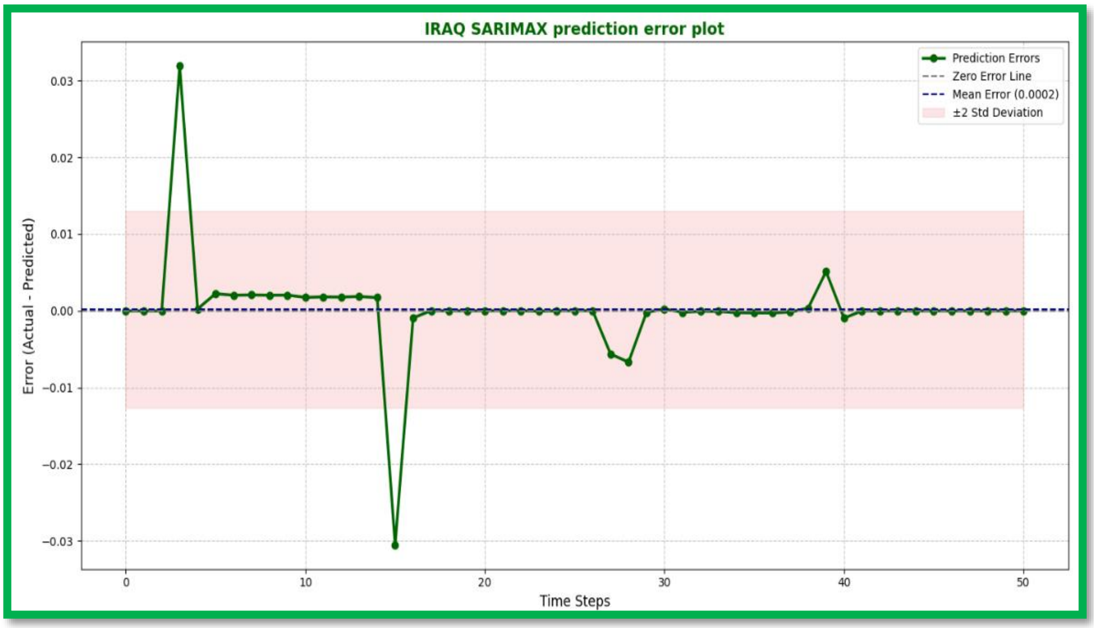

Figure 9. Iraq SARIMAX prediction error plot

According to the graph SARIMAX model's prediction errors for Iraq's EGDI are centered around zero. The mean error is a very low 0.0004 and the majority of the values fall within ±2 standard deviations in purple color. There are some initial peaks and troughs but these quickly stabilize. This indicates that the model does not suffer from time bias or structural skew, and confirms the quality of the predictions and the stability of the residuals.

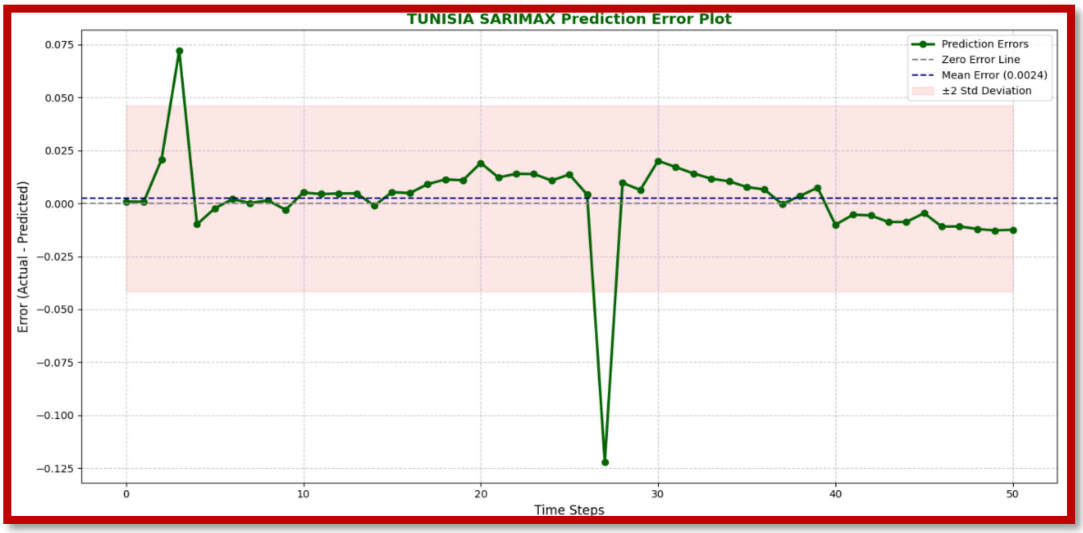

Figure 10. Tunisia SARIMAX prediction error plot

Figure 10 shows the EGDI for Tunisia prediction error high model reliability across the forecast horizon. Residuals remain consistently centered around the baseline without error, with a mean error estimate of approximately 0.0027. The majority of the forecast errors fall within the ±2 standard deviation confidence band and are highlighted in pink which indicating that the errors are statistically well distributed and free from systematic bias. One notable peak and one trough appear within the range they remain isolated and do not indicate a persistent structural anomaly. The error variance is low and generally stable and confirming that the model effectively captures the dynamics of the EGDI series.

5.2.1.2 A prediction error plot Test for Hybrid ARIMA-LSTM Model

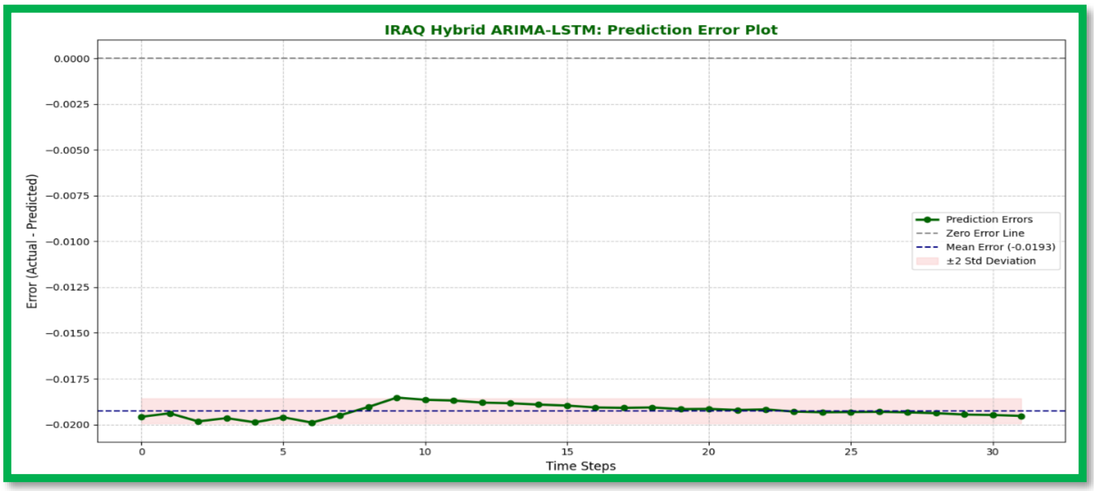

Figure 11. Iraq Hybrid ARIMA-LSTM prediction error plot

Figure 11 above illustrates the prediction error the actual minus the predicted for the proposed Iraq Hybrid ARIMA-LSTM model over the test period. Mean error as a horizontal dashed blue line is shown at approximately −0.0193, indicating that the model tends to slightly overestimate the actual values on average. Zero Error Reference Line: The grey dashed line at 0 represents the ideal prediction condition. The closeness of most points to this line supports the model's high precision. ±2 Standard Deviation The band of pink-shaded region indicates the confidence interval (±2σ) around the mean error. This band contains all of the points, indicating low dispersion and error stationarity. The deviation of the error stability curve is constant and of low magnitude. Additionally, indicating no notable outliers and temporal stability. Furthermore, throughout the forecasting horizon, the model retains its high accuracy and low bias. Strong generalization performance is indicated by the narrow confidence band and clustered errors around the mean. This further validates the adequacy of the model by conforming to the residual independence as confirmed by the Ljung-Box test.

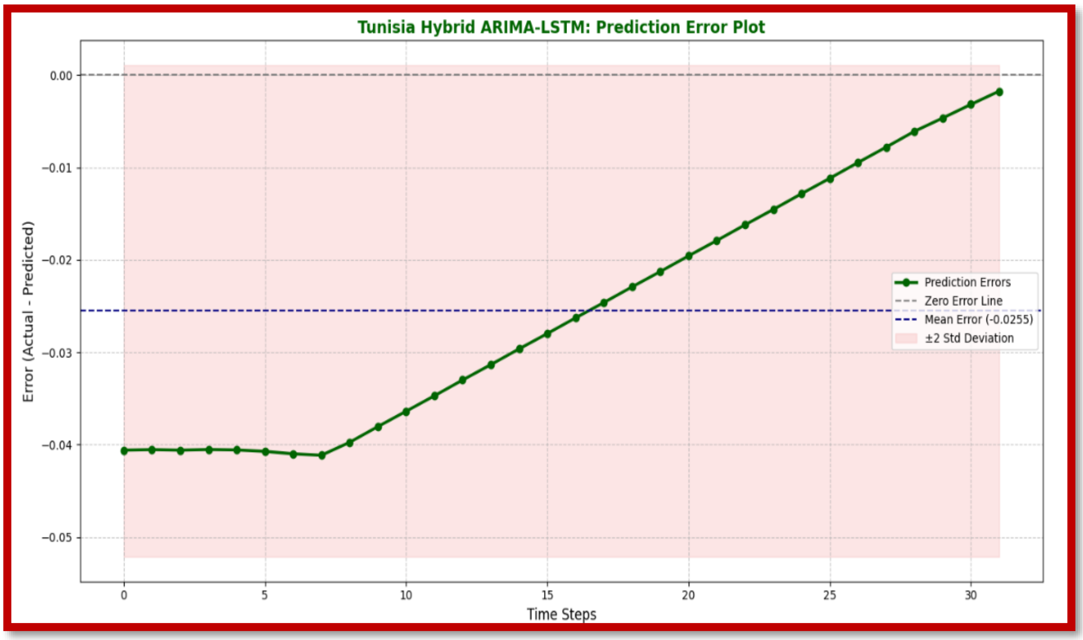

Figure 12. Tunisia Hybrid ARIMA-LSTM prediction error plot

The plot shows that all residual values fall within ±2 of the standard deviation, which statistically indicates the absence of outliers or sharp deviations in the forecast. This supports the hypothesis that the probability distribution of errors follows a stable normal pattern. The mean error as measured quantitatively is roughly -0.0255 a comparatively small value that shows no discernible bias in either direction (overestimation or underestimation) in the model. Furthermore, the trend of errors across time steps does not show a persistent systematic deviation or a cyclical pattern, confirming the independence of errors over time, a property that is necessary for the validity of time series models' fundamental assumptions. Based on these results, it can be concluded that the used hybrid ARIMA-LSTM model for demonstrates high efficiency in representing the structural and probabilistic properties of the EGDI data series for Tunisia, and supports its use in short- and medium-term forecasting.

5.3 Distribution Histogram, (KDE) Test

5.3.1 Distribution Histogram, (KDE) Test for SARIMAX Model

The tool evaluates a forecasting model's performance by comparing actual and predicted values over time, with errors rep-resented on the vertical axis and time intervals on the horizontal axis. It also uses prediction error distribution analysis and KDE-Kernel Density Estimation.

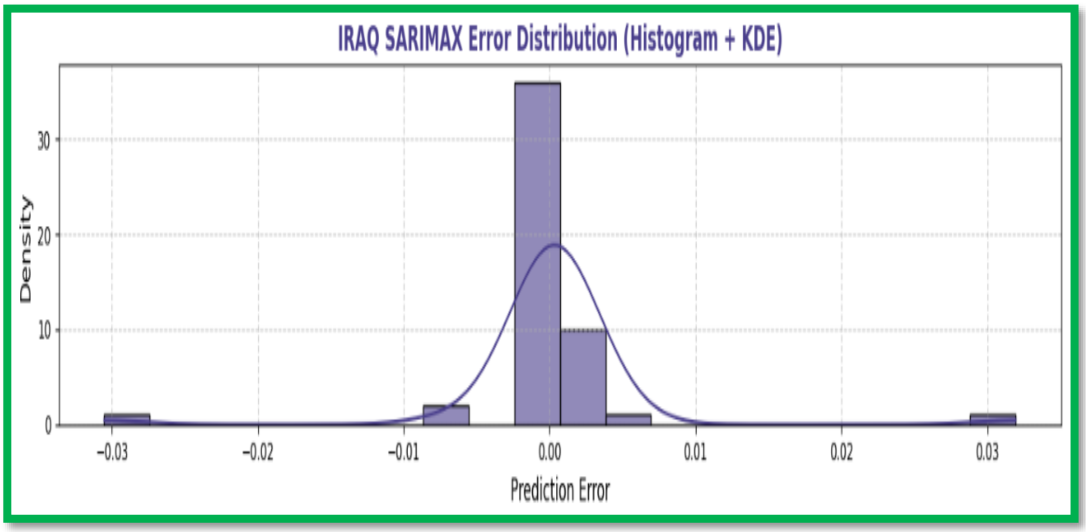

Figure 13. Iraq Error Distribution using a Histogram (KDE) Test for SARIMAX Model

Fig13 above clear the distribution of prediction errors for the SARIMAX model for the Iraq EGDI dataset is shown in the graph above. The errors are centered on zero. Since the projected values for the great majority of time, points closely resemble the actual values and the model's prediction accuracy is good. The regression model's assumptions are supported by the error distribution's proximity to a normal distribution. The residuals exhibit random behavior and lack systematic bias. The tail's few repetitions, which stand for comparatively greater faults, are uncommon and have little bearing on the model's quality. Additionally, the distribution's centrality shows that the majority of mistakes fall inside a small range. The model is appropriate for describing time-series data as this investigation shows.

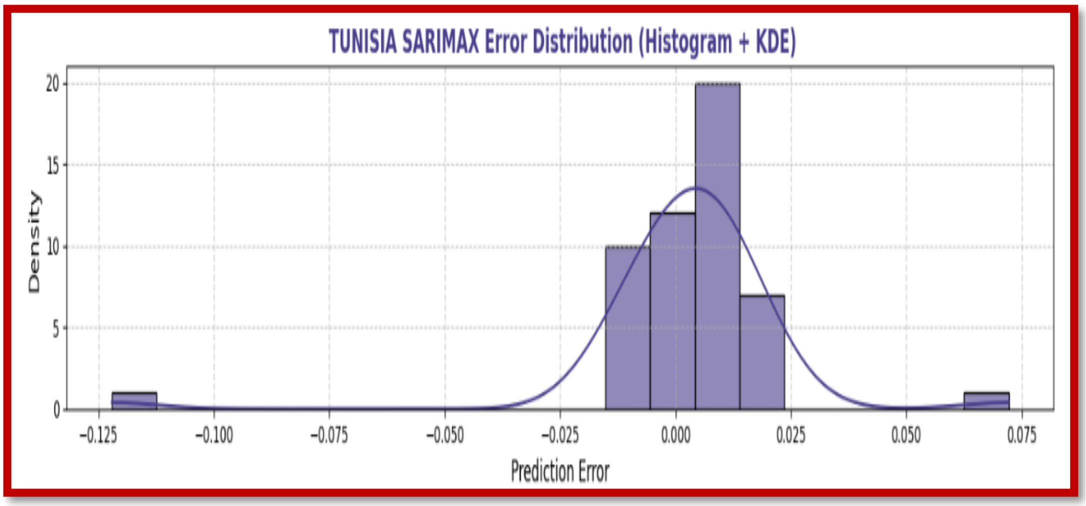

Figure 14. Tunisia Error Distribution using a Histogram (KDE) Test for SARIMAX Model

The Figure 14 the SARIMAX model error distribution plot for Tunisia shows a symmetric, unimodal structure centered near zero, indicating that the prediction errors are balanced without any evidence of significant bias. The frequency plot, along with the kernel density estimate, shows that the majority of errors are tightly clustered around the origin, reflecting high forecast accuracy and consistency. The bell-shaped curve is highly close to a normal distribution, meaning that the residuals behave like white noise. There are no heavy tails or multimodal peaks, indicating no structural distortions or model misspecifications. The smooth curvature and centered error density provide confirmation that the SARIMAX model adequately captures the underlying temporal dynamics of the EGDI data and maintains generalizability across observations.

5.4.2 Distribution Histogram, (KDE) Test for Hybrid ARIMA-LSTM Model

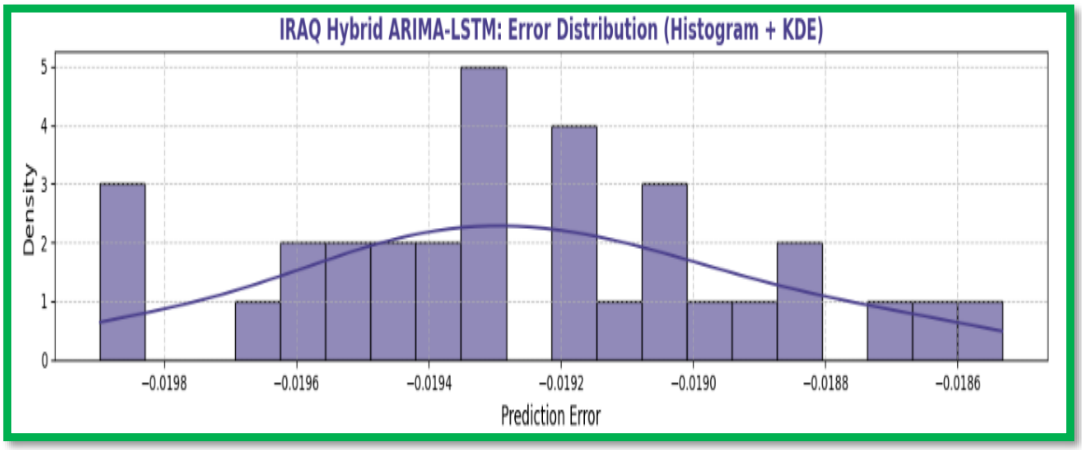

Figure 15. Iraq Error Distribution using a Histogram (KDE) Test for Hybrid ARIMA-LSTM Model

Forecast errors in the hybrid ARIMA-LSTM model for Iraq. It is observed that the distribution is centered on the annual child count value of -0.0193. The shape of the distribution is abruptly skewed outward, with no multiple peaks or skewed slopes, which is assumed a reason for a reasonably normal distribution. This result is significant for the current model in terms of forecast inconsistency. It strongly suggests that the strategic model achieved the desired results in capturing the data over the period without taking any significant action or significant risk in forecasting.

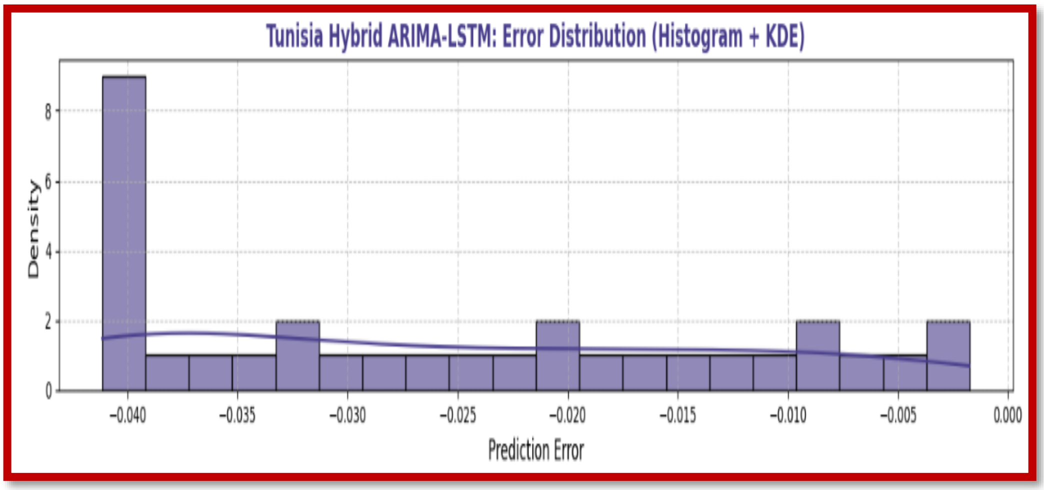

Figure 16. Tunisia Error Distribution using a Histogram (KDE) Test for Hybrid ARIMA-LSTM

The figure 16 shows the distribution of forecast errors generated by the hybrid ARIMA-LSTM model for Tunisia. Most of the error values are clustered around -0.04, due to a slight bias toward underestimating the actual values. Although the distribution does not follow a perfectly symmetrical pattern (asymmetric), the fact that most errors are clustered within a relatively narrow range indicates relative stability in the model's performance. There is a slight deviation from a standard normal distribution because the presence of cyclical structures in the data or some structural features not fully captured by the model. However, the distribution does not contain extreme peaks or sharp divergences and reinforcing the hypothesis that there are no errors affecting the forecast. Therefore, it can be concluded that the used ARIMA-LSTM model exhibits acceptable accuracy and a controlled error structure in representing the overall trend or minor seasonal components, and that there are no significant errors that would significantly impact its suitability for predicting the behavior of the EGDI series for Tunisia.

5.5 Time Series Decomposition (Additive) Test

A statistical technique used to analyze time series by separating them into their main components trend, seasonality, and residual. Additive decomposition is assumed that the time series can be expressed as the sum of these three components as follows: Y_t=T_t+S_t+R_t where Y_t Represent Total value of the series over time, T_t Trend over time, S_t Seasonal component and R_t Random component (residuals).

5.5.1 SARIMAX Time Series Decomposition Test

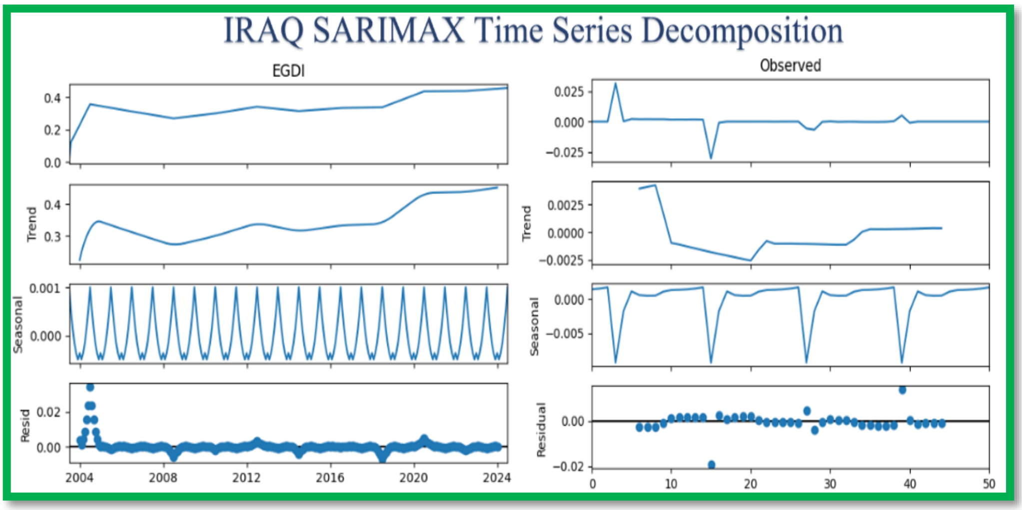

Figure 17. Iraq SARIMAX decomposition TEST

The figure 17 shows the analysis of the decomposition of the time series of the EGDI index in Iraq using the SARIMAX model. Left part represents the original series and shows an upward trend and clear annual seasonality with small regular residuals. Right part shows the analysis of the model's residuals, which appear to be, distributed around zero without clear patterns indicating the accuracy of the model and its success in capturing the time structure of the series.

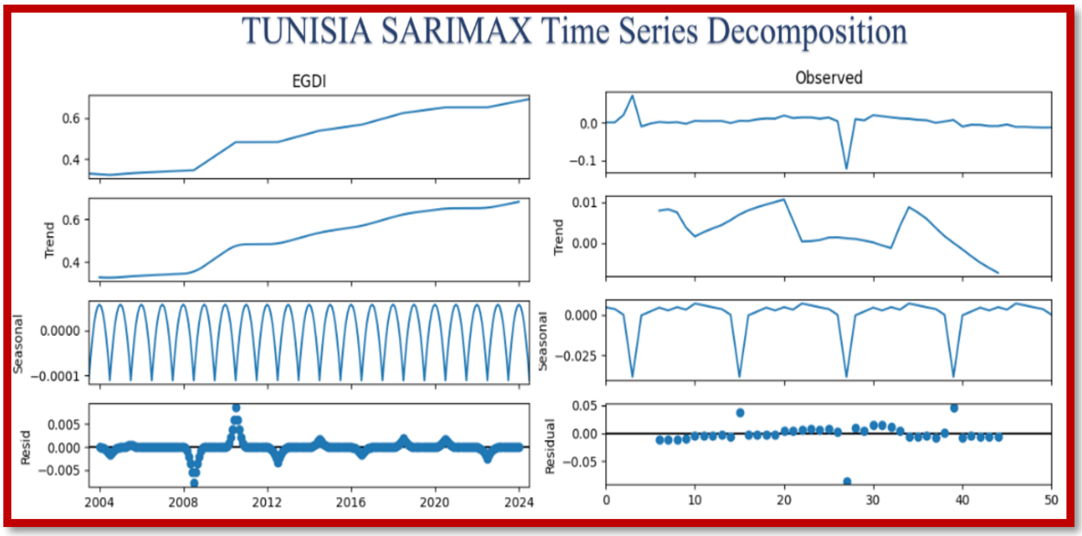

Figure 18. Tunisia SARIMAX decomposition TEST

The SARIMAX model was used to break down the EGDI series in Tunisia, as shown in Figure 18. The original data show a clear seasonality with low amplitude and a regular upward trend. The original residuals are small and move around a little. When looking at the results of the forecast residuals, they show a distribution around zero with a few outliers but no clear pattern. The model is good and can accurately show how the series changes over time.

5.5.2 Hybrid ARIMA-LSTM Time Series Decomposition Test

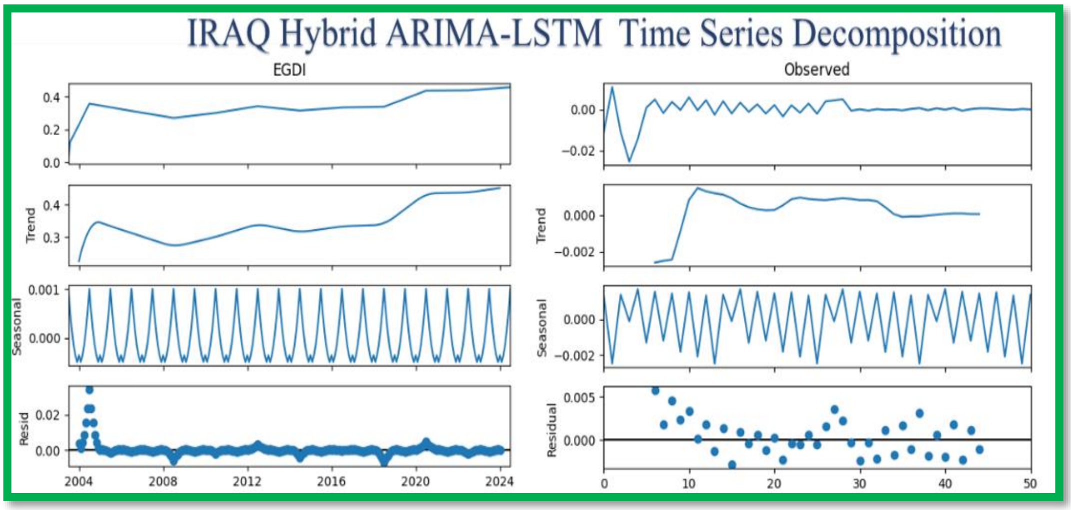

Figure 19. Iraq Hybrid ARIMA-LSTM decomposition TEST

Using a combination of ARIMA and LSTM models, this graph looks at the time series of the EGDI for Iraq. There was a steady rise over time, with a clear seasonal pattern throughout the observed period. The residuals of the original data are still close to zero, which means that the model fits the data well. From a forecasting point of view, there are very small changes at first, but the results quickly become stable. As it goes along, the model changes. There is a seasonal pattern in the errors in the forecasts, but it does not have a big impact on how accurate they are. Overall, the model does a good job of following the data structure, with only small changes at the start of the forecast window. The model can find both general and seasonal errors, which makes it a reliable model.

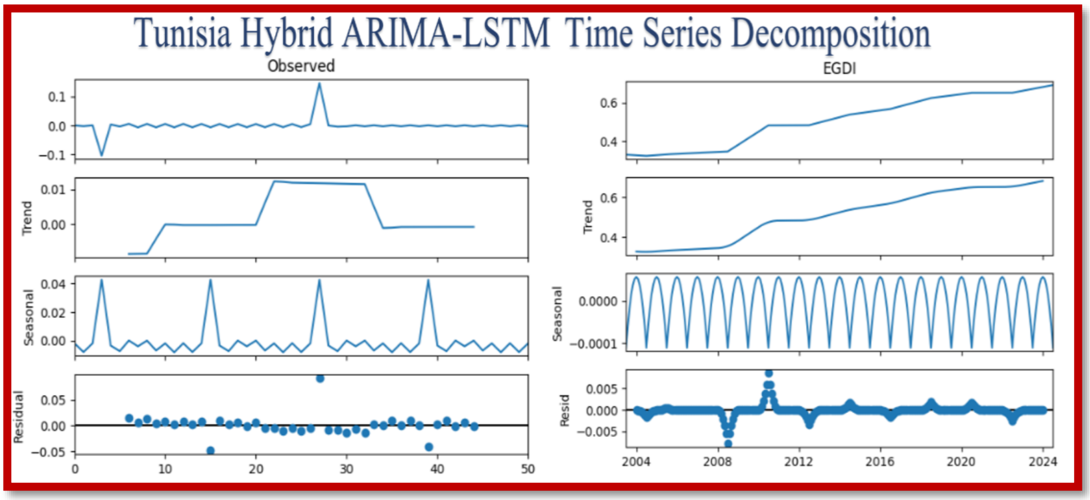

Figure 20. Tunisia Hybrid ARIMA-LSTM decomposition TEST

The decomposition of Tunisia's EGDI series utilizing the hybrid ARIMA-LSTM model provides a detailed analysis of the data's temporal behavior. Original EGDI series shows a steady upward trend with a strong and consistent seasonal pattern. The residuals from the original data remain small and centered around zero that mean model fits the historical values well. Looking at the forecast components on the left side, the observed prediction errors start with minor fluctuations and then stabilize. The trend in the forecast remains generally flat, while the seasonal component appears more irregular compared to the original series, but still manageable. Most of the prediction residuals are tightly distributed with no major outliers, suggesting that the hybrid model captures both structure and dynamics of the data reasonably well. There is slight variation in the seasonal behavior during prediction. Overall, the model manages to handle both trend and seasonality effectively in Tunisia's EGDI data.

5.6 TEST 3: Model Evaluation Metrics

Model evaluation metrics are statistical standards that are used to assess how effectively forecasting models perform and how closely they match actual values. These seven measures aid in evaluating various models and selecting the most effective one according to forecast accuracy.

5.6.1 Model evaluation metrics Test for SARIMAX Model

Table 7: IRAQ and Tunisia SARIMAX Model Evaluation Metrics Test

Evaluation Metrics of EGDI SARIMAX Model IRAQ and TUNISIA			
Metric	IRAQ	TUNISIA	Model Performance Comparison
RMSE (Root Mean Square Error)	0.0063	0.0228	IRAQ Lower Error
MAE (Mean Absolute Error)	0.0019	0.0119	IRAQ Lower Error
MAPE (Mean Absolute Percentage Error)	0.0165	0.0207	IRAQ Lower Relative Error
MAPA (Prediction Accuracy %)	98.35%	97.93%	IRAQ Higher Accuracy
AIC (Akaike Information Criterion)	-1682.933	-677.262	Iraq -Better Model Complexity & Accuracy Tradeoff
BIC (Bayesian Information Criterion)	-1662.437	-639.832	Iraq -Better Model Complexity & Accuracy Tradeoff
MHE (Mean Hellinger Distance)	0.0378	0.0278	TUNISIA Better Similarity Between Actual & Predicted Distribution

Table 7 represents a detailed comparison of the SARIMAX model's performance evaluation indicators for EGDI prediction in both Iraq and Tunisia. The results show that the model in Iraq performed better on most indicators. The lower RMSE and MAE values in Iraq indicate lower absolute and root mean prediction errors, reflecting higher accuracy. The MAPE value in Iraq was 0.0165 compared to 0.0207 in Tunisia, indicating a relatively lower error rate. The prediction accuracy (MAPA) in Iraq reached 98.35% compared to 97.93% in Tunisia, confirming the model's superiority in terms of accurate predictions. On the other hand, the AIC and BIC values in Iraq were significantly lower, indicating a better balance between model complexity and quality. As for the MHE, it was lower in Tunisia (0.0278) compared to Iraq (0.0378), indicating that the distribution of predictions in Tunisia was closer to the actual distribution of values. Although the Iraqi model outperformed in accuracy and lower error, the difference in MHE is attributed to the fact that the Tunisian data underwent double differential sampling, which affected the original distribution but helped the distributional shape of the predictions converge to the actual values. Accordingly, the Iraqi model can be considered more accurate and less complex, while the Tunisian model demonstrated better distributional convergence, albeit at the expense of some accuracy.

5.6.2 Model Evaluation Metrics Test for Hybrid ARIMA-LSTM Model

Table 8: Iraq and Tunisia Hybrid ARIMA-LSTM Model Evaluation Metrics Test

Evaluation Metrics of EGDI Hybrid ARIMA-LSTM Model Iraq and Tunisia

Metric	IRAQ	TUNISIA	Model Performance Comparison
RMSE (Root Mean Square Error)	0.0193	0.0288	IRAQ Lower Error
MAE (Mean Absolute Error)	0.0193	0.0255	IRAQ Lower Error
MAPE (Mean Absolute Percentage Error)	0.0432	0.0386	TUNISIA Lower Relative Error
MAPA (Prediction Accuracy %)	95.68%	96.14%	TUNISIA Higher Accuracy
AIC (Akaike Information Criterion)	-777.028	-771.282	Model results are close Complexity & Accuracy Tradeoff
BIC (Bayesian Information Criterion)	-755.9	-750.177	Model results are close Complexity & Accuracy Tradeoff
MHE (Mean Hellinger Distance)	0.0007	0.0141	IRAQ Better Similarity Between Actual & Predicted Distribution

The table 8 shows a detailed comparison of the results of the Hybrid ARIMA-LSTM model for predicting the EGDI index in Iraq and Tunisia based on various evaluation indicators. The model's performance in terms of RMSE and MAE is lower in Iraq (0.0193), indicating that the Iraqi model provided more accurate predictions in absolute terms. This mean good numerical stability in the model's results in Iraq compared to Tunisia. The error and precision (MAPE) are lower in Tunisia (0.0386) meaning that the error relative to the actual values was better in Tunisia. MAPA accuracy is also higher in Tunisia (96.14%) compared to 95.68% in Iraq, indicating that the Tunisian model was more accurate in the percentage of correct predictions. We conclude that Tunisia outperforms Iraq in relative accuracy. The AIC and BIC values are close between Iraq and Tunisia, but Iraq has slightly lower values (larger and negative), indicating lower complexity or slightly higher efficiency in the Iraqi model. The explanation attached to the table confirms that the difference is not significant and that the results are similar in this aspect. The MHE (Hellinger Distance) value in Iraq is significantly lower (0.0007) than in Tunisia (0.0141), indicating that the statistical distribution of Iraq's forecasts is much closer to the actual distribution. This is a strong indication that the Iraq model is more consistent in terms of its forecast structure than the actual values. The Tunisia model showed higher relative accuracy (MAPA) and MAPE making it good at predicting values as a percentage of the original value, but it was less accurate in terms of absolute error and distribution. The Iraq model outperformed in terms of reducing the direct error and converging to the statistical distribution, with a slight advantage in model complexity according to the AIC and BIC. Thus, the Tunisia model can be considered stronger in relative forecasts, and the Iraq model is more stable and numerically accurate in distribution and absolute errors.

6. Models Forecasting Results

In this section presents the results of applying four predictive models to the EGDI series for Iraq and Tunisia. Two conventional SARIMAX models and two advanced Hybrid ARIMA-LSTM models were constructed. The models were built to evaluate the prediction accuracy and time series representation efficiency for each country. The proposed approach to construct these models relied on 12 visual tests and statistical performance indicators such as RMSE, MAE, MAPE, MHE, AIC, BIC, and MAPA (Accuracy). The results of the following four models reflect the subtle differences in model performance between the two countries and between conventional and hybrid models.

6.1 IRAQ SARIMAX Predictive Model

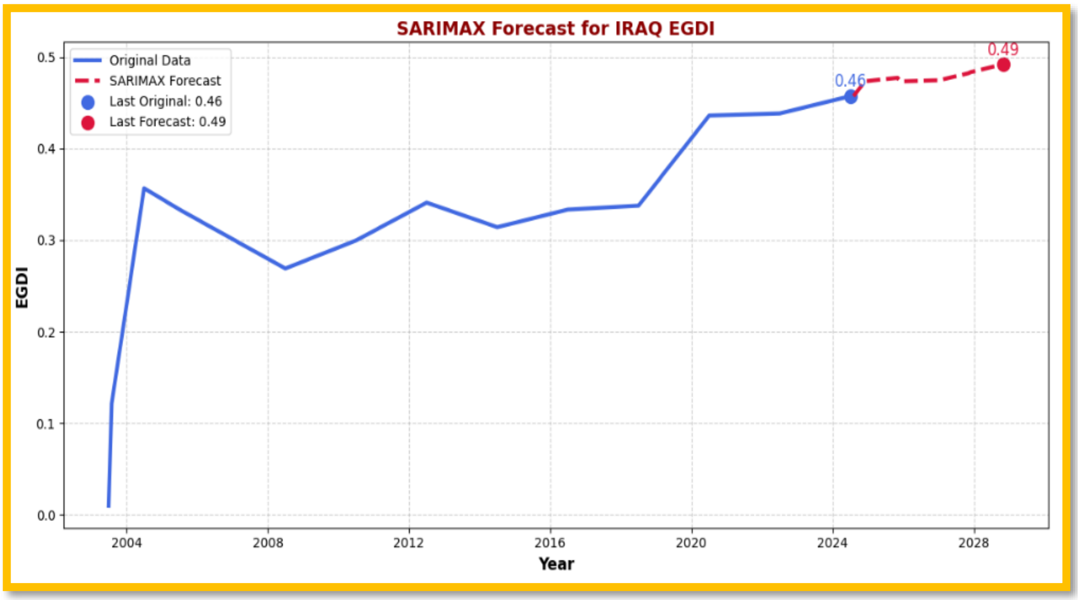

Figure 21. Iraq SARIMAX Predictive Model

The figure 21 shows the projected growth of the Iraq E-Government Development Index (EGDI) using the SARIMAX model. The solid blue line represents the original actual data, while the dashed red line represents the model's future projections starting from the year following the last year of actual data and continuing upward. The model predicts positive expectations in growth of the Iraq EGDI over the coming years. The blue dot represents the last recorded actual value of the index, which is 0.46, while the second red dot indicates the projected value after four and a half years, which is 0.49. The figure demonstrates the model's high ability to keep pace with the historical trends of the index while maintaining continuous growth in future projections.

6.2 Tunisia SARIMAX predictive model

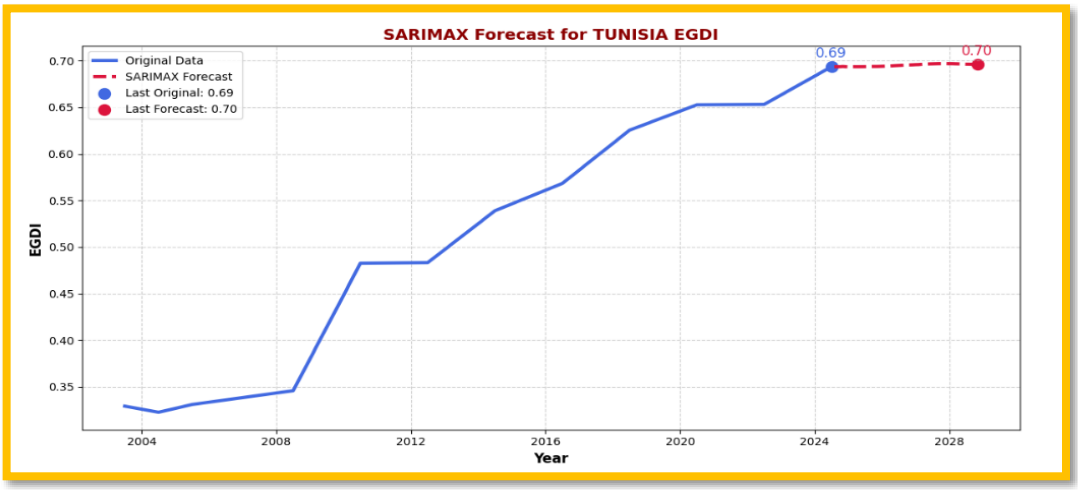

Figure 22. Tunisia SARIMAX Predictive Model

Figure 22 shows SARIMAX forecast chart for Tunisia's EGDI shows a steady upward trajectory in historical data. The model successfully captures this trend and extends it, with a modest increase expected for the coming years. The forecast line predicts this linear growth to continue, with stability and slight improvement in e-government performance. SARIMAX model predicts a small change between the first observed value of 0.69 and the last predicted value of 0.7. The model has a strong fit and minimal discontinuity, which reinforces the validity of the SARIMAX construct. The smooth transition from the training phase to the prediction phase reflects the model's ability to generalize from past patterns without overfitting.

6.3 IRAQ HYBRID ARIMA -LSTM predictive model

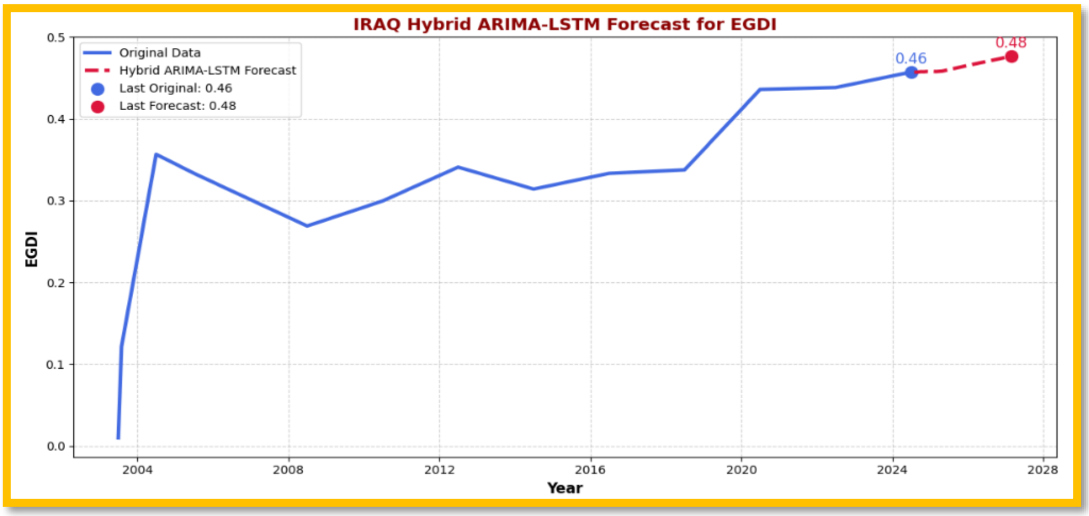

Figure 23. Iraq Hybrid ARIMA-LSTM Predictive Model

The ARIMA-LSTM hybrid forecasting model was used to make predictions about the Electronic Government Development Index (EGDI) for Iraq. The blue graph shows the original time series of EGDI values from 2003 to the last observation point, which was in 2024 and had a value of 0.46. The dashed red lines show what is expected to happen at the end of 2028, with the last value being 0.48. The graph shows that the model is able to accurately predict the long-term rise in the data and that EGDI is likely to continue to improve slightly in the next few years. The blue and red lines cross at the turning point, which means that the previous time series will keep going without any sudden changes. This shows that the model is very re-liable for short-term extrapolation.

6.4 TUNISIA HYBRID ARIMA -LSTM Predictive Model

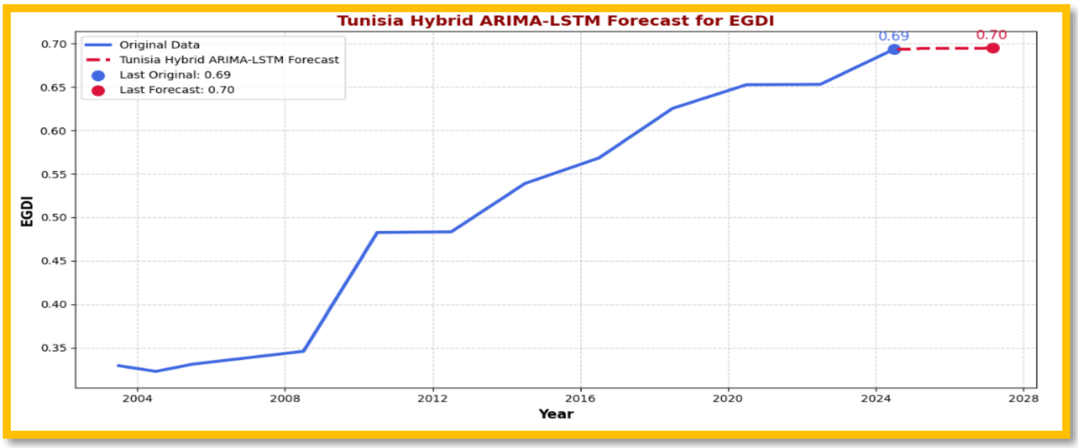

Figure 24. Tunisia Hybrid ARIMA-LSTM Predictive Model

The figure 24 above shows forecasts using the hybrid ARIMA-LSTM model for the E-Government Development Index (EGDI) in Tunisia. We observe the last actual value of 0.69 in 2024. The hybrid model projects the index to rise to 0.70 in the coming years, driven by e-government development in Tunisia. This continued upward trend in the actual data supports the hypothesis of sustainable growth and enhances confidence in the model's predictive performance. The absence of sharp fluctuations in future forecasts also indicates the model's stability and its lack of impact from seasonal fluctuations or recent structural shocks.

7. Discussion

Traditional SARIMAX models outperformed through the Iraqi model achieving the highest predictive accuracy of 98.35% along with the lowest RMSE (0.0063) and MAE (0.0019). The AIC and BIC indices recorded lower values in Iraq. Besides, reflecting a more efficient model in balancing accuracy and complexity. In contrast, the Tunisian SARIMAX model demonstrated slightly lower predictive accuracy of 97.93%. However, was distinguished by better distributional similarity between predicted and actual values, according to the MHE index (0.0278 versus

0.0378), demonstrating the Tunisian model's ability to reproduce the shape of statistical series with greater accuracy. In advanced Hybrid ARIMA-LSTM models the Tunisian model performed better in terms of relative accuracy (96.14%) compared to the Iraqi model (95.68%) with a significant improvement in MAPE (3.86%) versus (4.32%). Which means the effectiveness of LSTMs in learning and correcting residual errors in the Tunisian series. In contrast, although the hybrid model in Iraq did not outperform in terms of MAPA, it had the highest distributional fit to the true values, recording the lowest MHE of all (0.0007), demonstrating the accuracy of reconstructing the overall behavior of the series. When comparing each country with its two models, the SARIMAX model in Iraq clearly outperforms the hybrid model in terms of accuracy (98.35% versus 95.68%), while Tunisia achieves better performance using the hybrid model (96.14% versus 97.93%), although the difference is small. The most important finding in our research that the effectiveness of the model depends on the characteristics of the time series itself. The Iraqi series that exhibits season-al stability and a clear trend is better suited to linear models such as SARIMAX while the Tunisian series with its more di-verse dynamic patterns benefits from LSTM's ability to model nonlinear relationships.

8. Conclusion

The results of evaluating the four models using multiple indices, most notably the MAPA, show that performance varies significantly depending on the characteristics of each EGDI time series and the type of model used. The SARIMAX model for Iraq performed best, achieving the highest predictive accuracy among all models (MAPA = 98.35%), along with the lowest error rates (RMSE and MAE), and the best values for the AIC and BIC indices. Because of, the trend stability and clear seasonality of the Iraqi time series. These characteristics make the SARIMAX model an ideal choice, given its reliance on linear and seasonal components. Hybrid ARIMA-LSTM model for Tunisia performed better to achieving higher accuracy (MAPA = 96.14%) and a slightly higher MHE which was still within an acceptable range. Hybrid ARIMA-LSTM model ability to handle the more dynamic and complex changes in the Tunisian series as the LSTM layer provided additional flexibility in capturing nonlinear relationships and correcting residual errors after applying ARIMA. Based on the above, SARIMAX is the optimal choice in cases where the time series is relatively stable and regular in its trend and seasonal pattern as is the case in Iraq. However, in cases containing nonlinear elements and sudden fluctuations or irregular temporal behavior hybrid models such as Hybrid ARIMA-LSTM offer better performance as cleared in the Tunisian data, thanks to their ability to adaptively learn and correct subtle temporal errors.

9. Recommendations and Future Directions

In the light of findings, it is recommend conducting this 12-test analysis approach for evaluating future predictive models. This comprehensive analytical coverage balances the accuracy and ease the understanding uncovering the hidden model shortcomings. This methodology includes three integrated stages.

1. First, Residuals Analysis (ACF), PACF, and Ljung-Box.
2. Second, Prediction Visualization and Distribution contain a Prediction Error Plot, a Distribution Histogram, (KDE) analysis of errors, and Time Series Decomposition (Additive) analysis.
3. Third, Quantitative Evaluation (QE), MAE, MAPE, RMSE, AIC, BIC, MHE, and MAPA.

This methodology is integrated theoretical coverage and practical application. It allows for a comprehensive evaluation of the model from statistical, graphical, and distributional aspects, and provides a high ability to detect weaknesses or biases in predictive models. It has proven its effectiveness in this research on classical predictive models (SARIMAX) and advanced hybrid models (Hybrid ARIMA-LSTM).

The 12 tests ensuring that the predictive model not only minimizes error but also preserves the underlying statistical structure of the time series. Collectively, these indicators formed a robust framework for evaluating and comparing the predictive performance of SARIMAX and Hybrid ARIMA-LSTM models across different datasets. Accordingly, we recommend adopting this evaluation framework as a standard methodology in future studies dealing with analysis and prediction. Time series, because they provide comprehensiveness, accuracy, and flexibility in evaluation, regardless of the type of model or field of application.

References

[1] United Nations, "E-Government Survey 2004."

[2] U. UNDESA, "Government Survey 2010: Leveraging e-government at a time of financial and economic crisis," Available: www2.unpan.org/egovkb/global_reports/10report.htm, 2010.

[3] D. of E. and S. Affairs, United Nations E-Government Survey 2024: Accelerating Digital Transformation for Sustainable Development-with the Addendum on Artificial Intelligence. United Nations Research Institute for Social Development, 2024.

[4] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.

[5] W. F. H. Zhang and X. Zheng, "A hybrid ARIMA and LSTM model in time series forecasting," *Procedia Computer Science*, vol. 162, pp. 103–109, 2019.

[6] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2017.

[7] L. Zhang, W. Liu, and B. Qi, "Energy optimization of multi-mode coupling drive plug-in hybrid electric vehicles based on speed prediction," *Energy*, vol. 206, p. 118126, 2020.

[8] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?," *Journal of Econometrics*, vol. 54, no. 1–3, pp. 159–178, 1992.

[9] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366a, pp. 427–431, 1979.

[10] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. Springer, 2002.

[11] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 2003.

[12] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, pp. 461–464, 1978.

[13] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2018.

[14] R. H. Shumway, D. S. Stoffer, and D. S. Stoffer, *Time Series Analysis and Its Applications*, vol. 3. Springer, 2000.

[15] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica: Journal of the Econometric Society*, pp. 987–1007, 1982.

[16] R. S. Tsay, *Analysis of Financial Time Series*. John Wiley & Sons, 2005.

[17] J. H. Stock and M. W. Watson, *Introduction to Econometrics*. Pearson, 2020.

[18] W. Enders, "Applied Econometric Time Series," *Privredna Kretanja i Ekonomska Politika*, vol. 132, p. 93, 2012.

[19] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences*, vol. 100. Academic Press, 2011.

[20] T. Gneiting and M. Katzfuss, "Probabilistic forecasting," *Annual Review of Statistics and Its Application*, vol. 1, no. 1, pp. 125–151, 2014.

[21] C. Chatfield, *Time-Series Forecasting*. Chapman and Hall/CRC, 2000.

[22] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[23] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting," *Applied Soft Computing*, vol. 11, no. 2, pp. 2664–2675, 2011.

[24] F. R. Alharbi and D. Csala, "A seasonal autoregressive integrated moving average with exogenous factors (SARIMAX) forecasting model-based time series approach," *Inventions*, vol. 7, no. 4, p. 94, 2022.

[25] Statsmodels.org, "SARIMAX Model." [Online]. Available: https://www.statsmodels.org/dev/examples/notebooks/generated/statespace_sarimax_stata.html. [Accessed: May 19, 2025].

[26] N. S. Arunraj, D. Ahrens, and M. Fernandes, "Application of SARIMAX model to forecast daily sales in food retail industry," *International Journal of Operations Research and Information Systems (IJORIS)*, vol. 7, no. 2, pp. 1–21, 2016.

[27] U. Nations, "E-Government Survey 2005: From E-Government to E-Inclusion."

[28] U. Nations, "E-Government Survey 2022: The Future of Digital Government," United Nations Department of Economic and Social Affairs, 2022.

[29] U. Nations, "E-Government Survey 2020: Digital Government in the Decade of Action for Sustainable Development," Department of Economic and Social Affairs, United Nations, New York, 2020.

[30] U. N. E.-G. Survey, “Gearing e-government to support transformation towards sustainable and resilient societies,” 2018.

[31] U. Nations, “E-Government Survey 2016: E-Government for Sustainable Development,” 2016.

[32] V. Nationen, *United Nations E-Government Survey 2014: E-Government for the Future We Want*. UN, 2014.

[33] United Nations, “E-Government Survey 2012: E-Government for the People,” 2012.

[34] U. Nations, “UN E-Government Survey 2008: From E-Government to Connected Governance,” UN Global E-Government Survey 2003. Department of Economic and Social Affairs, 2008.