Time Series Analysis and Forecasting of IDX Composite Index (IHSG) Based on ARIMA and ETS Models using Python

Forecasting the IDX Composite Index (IHSG) level with Time Series Analysis using two models, namely the ARIMA model and the ETS Model.

Mufad Luqman Nur Hakim

Project Overview

IDX Composite or Indeks Harga Saham Gabungan (IHSG)

According to the Indonesia Stock Exchange website, IDX Composite is an index that measures the stock price performance of all listed companies in Main Board and Development Board of the Indonesia Stock Exchange.

Objectives

- Forecasting the IDX Composite or Indeks Harga Saham Gabungan (IHSG) level for the following year.
- Determining the best performance in forecasting, in this project between the ARIMA model and the ETS Model.

Model

- ARIMA (Autoregressive Integrated Moving Average)
- ETS (Exponential Smoothing)

Dataset

The data consists of approximately **8,000 rows of daily Date, Open, High, and Close** data from the IDX Composite for the **period 1995-2025**. The data is sourced from Yahoo Finance and compiled by Gareth Aurelius Harrison. The data can be downloaded via this link: Kaggle

Note: This analysis is not intended as financial advice; it is for informational purposes only in the application of scientific studies that are expected to provide perspectives related to predictions or forecasts of the IDX Composite/Indeks Harga Saham Gabungan (IHSG) Level.

Workflow

Loading Data

Prepare the data set and import it into Jupyter Notebook in Google Colab

Data Preprocessing

Checking for missing values, duplicate data, and cleaning them up.

Resampling

Convert the data into a monthly form of the average for each month..

Exploratory Data Analysis

Raw data, checking stationarity, checking seasonality and trend.

Data Split (Test-Train)

Split the data into two parts, for test data and training data.

Create Model

Defining and Creating ARIMA Model and ETS Model.

Model Diagnostics

Diagnose residual models with visualizations, such as ACF, PACF, etc.

Model Evaluation

Execute and evaluate model performance with test data.

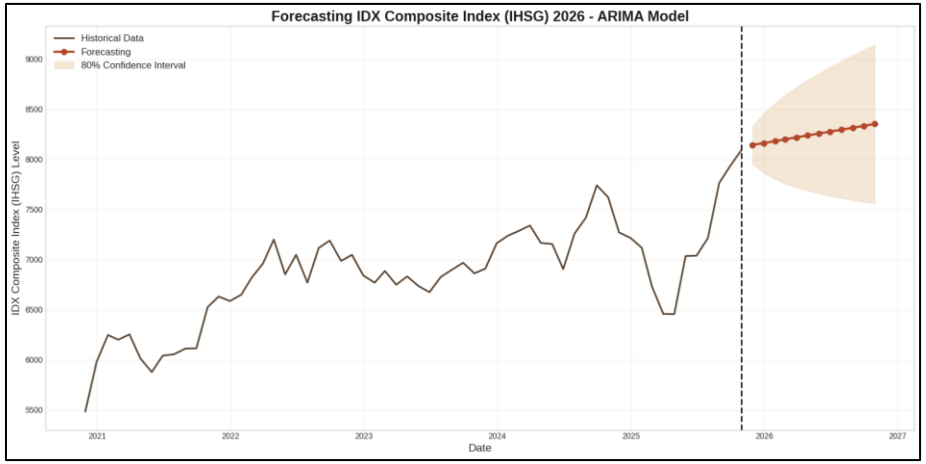
Forecasting

Performing forecasting with both models (ARIMA and ETS Model)

Conclusion

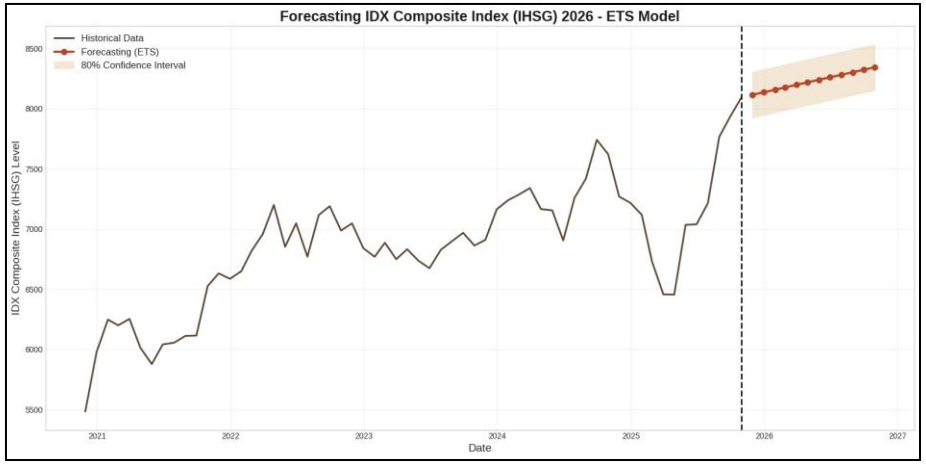
Get conclusions from model performance, including forecasting data.

Forecasting
IDX Composite Index (IHSG) 2026
ARIMA Model



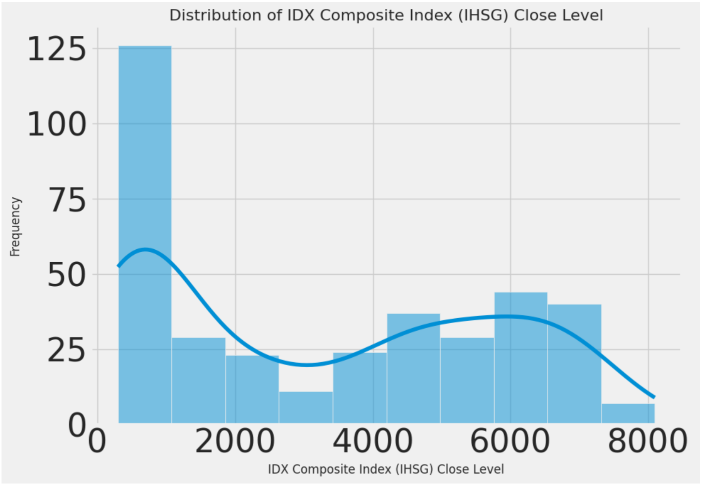
Data Forecasting ARIMA Model			
Date	Forecast	Lower 80%	Upper 80%
11/30/2025	8142.227367	7952.507419	8331.947315
12/31/2025	8161.27287	7861.213808	8461.331931
1/31/2026	8180.318372	7800.742452	8559.894292
2/28/2026	8199.363875	7754.256446	8644.471303
3/31/2026	8218.409377	7716.45127	8720.367484
4/30/2026	8237.454879	7684.990976	8789.918783
5/31/2026	8256.500382	7658.229491	8854.771273
6/30/2026	8275.545884	7635.092494	8915.999274
7/31/2026	8294.591387	7615.243781	8973.939015
8/31/2026	8313.636889	7598.002424	9029.271355
9/30/2026	8332.682391	7583.451554	9081.913229
10/31/2026	8351.727894	7570.713699	9145.742089

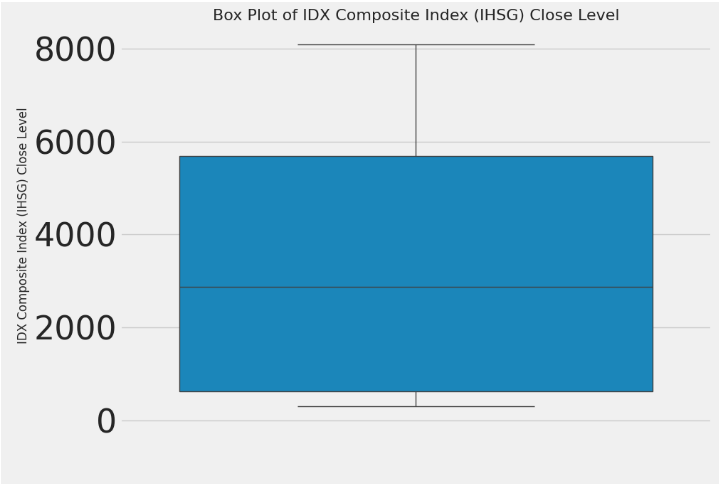
Forecasting
IDX Composite Index (IHSG) 2026
ETS Model

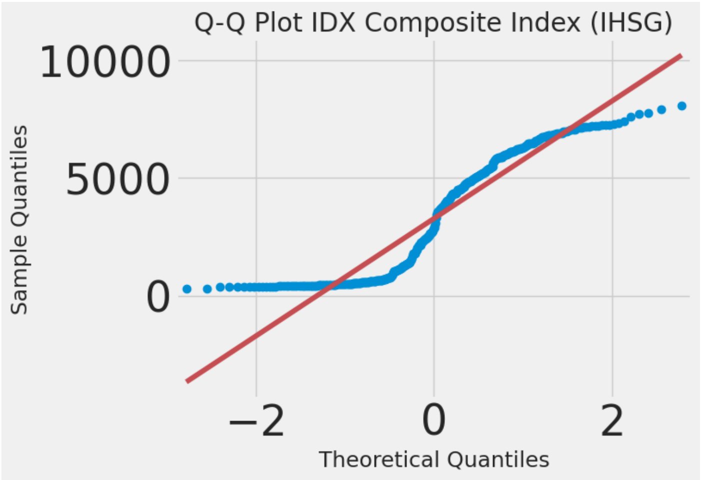


Data Forecasting ETS Model			
Date	Forecast	Lower 80%	Upper 80%
11/30/2025	8113.999	7923.391	8304.667
12/31/2025	8134.724	7944.116	8325.332
1/31/2026	8155.449	7964.84	8346.057
2/28/2026	8176.174	7985.566	8366.782
3/31/2026	8196.9	8006.291	8387.507
4/30/2026	8217.624	8027.016	8408.231
5/31/2026	8238.349	8047.74	8428.956
6/30/2026	8259.074	8068.465	8449.68
7/31/2026	8279.799	8089.19	8470.405
8/31/2026	8300.524	8109.915	8491.13
9/30/2026	8321.249	8130.64	8511.854
10/31/2026	8341.974	8151.366	8532.581

Data Distribution
Analyzing data distribution is very important because it will affect modeling, so data treatment must be considered in accordance with data distribution. Below is a visualization of data distribution using data distribution visualization, Box Plot, and Q-Q Plot.



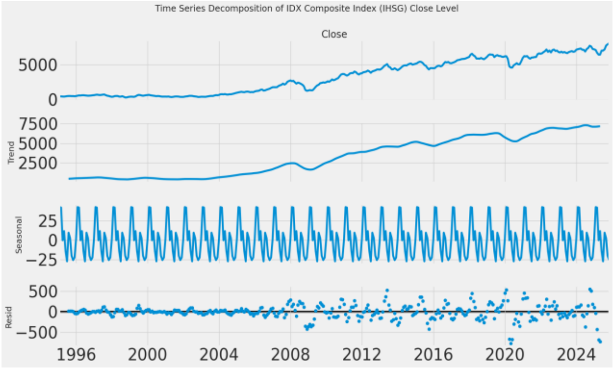




Distribution Plot Inference
Based on the data distribution visualization above, the data from IDX Composite is not normally distributed; the data appears to be right-skewed.

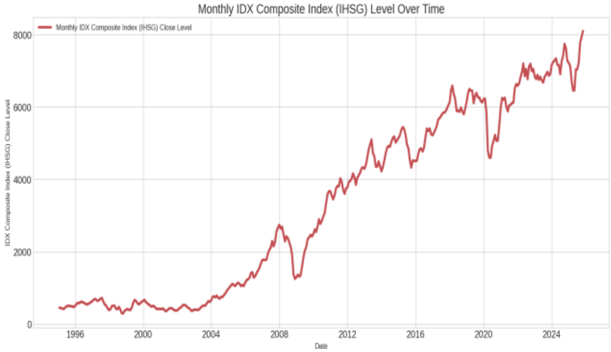
Box Plot Inference
Based on the visualization, the distribution shows a right skew, with outlier values approaching the upper limit around the 8000 level.

Q-Q Plot
Heavy Tail Distribution - The curve at the end shows the slope on the right and left sides, indicating that the distribution does not follow a normal Gaussian distribution.



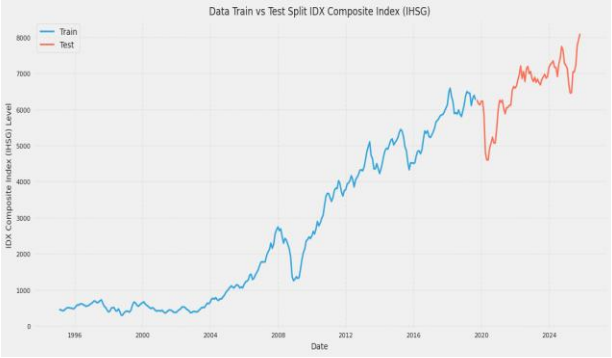
Decomposition

Decomposition is a statistical process that breaks down complex time series data into its basic components, such as trends, seasons, etc. Its main purpose is to understand the factors driving data movements and improve forecasting accuracy.
In the **Seasonal visualization**, seasonal patterns are visible but their strength is uncertain, indicating that this data is seasonal. In the **Trend visualization**, it can be seen that there is an overall increase, indicating that there is a trend in this data.



Resampling

The main purpose of resampling data into monthly frequencies is to simplify analysis, identify long-term trends, and reduce noise in the data. This process summarizes many data points from higher frequencies (such as daily) into one value per month. In this data, the values are daily, so they are converted into monthly data using averages.

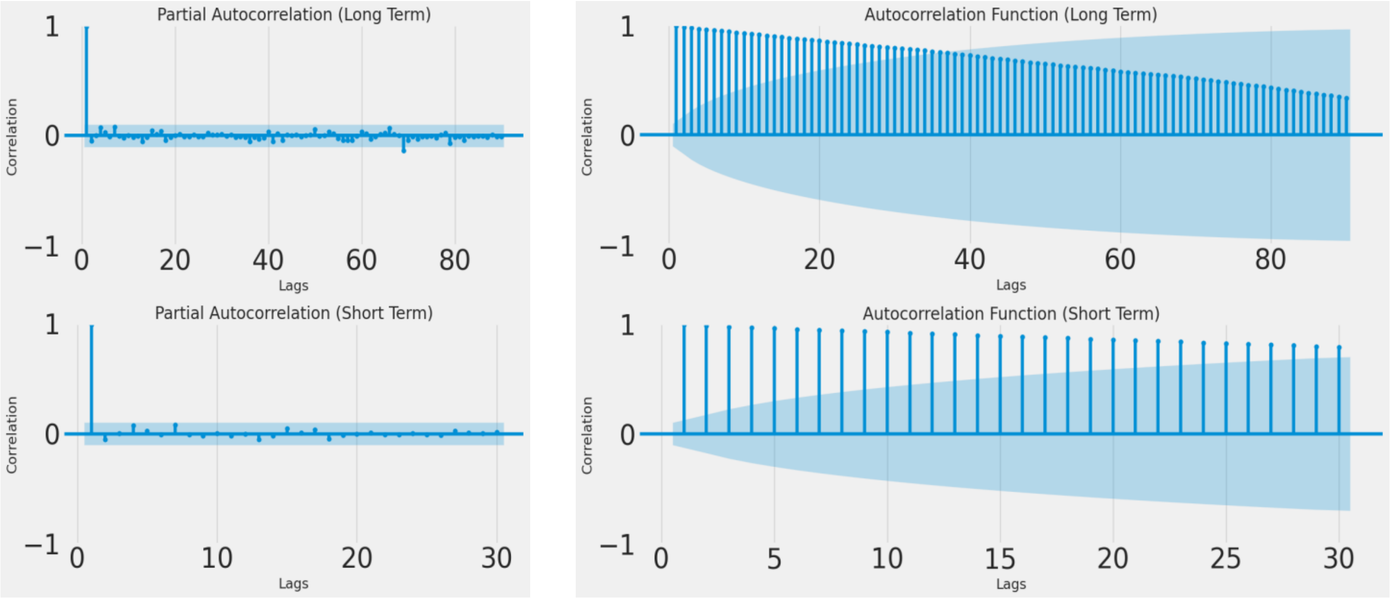


Data Splitting

Data splitting for time series is an important where sequential data sets are divided into several parts, usually for training, validation, and testing. The main difference with regular data is that in time series, the chronological order of the data must be preserved. The data cannot be randomized, as this would cause data leakage, where the model “sees” the future to predict the past.

Stationarity

Stationarity means that the statistical properties of a time series i.e. mean, variance and covariance do not change over time. Many statistical models require the series to be stationary to make effective and precise predictions. Two statistical tests would be used to check the stationarity of a time series – Augmented Dickey Fuller ("ADF") test and Kwiatkowski-Phillips-Schmidt-Shin ("KPSS") test. A method to convert a non-stationary time series into stationary series shall also be used.



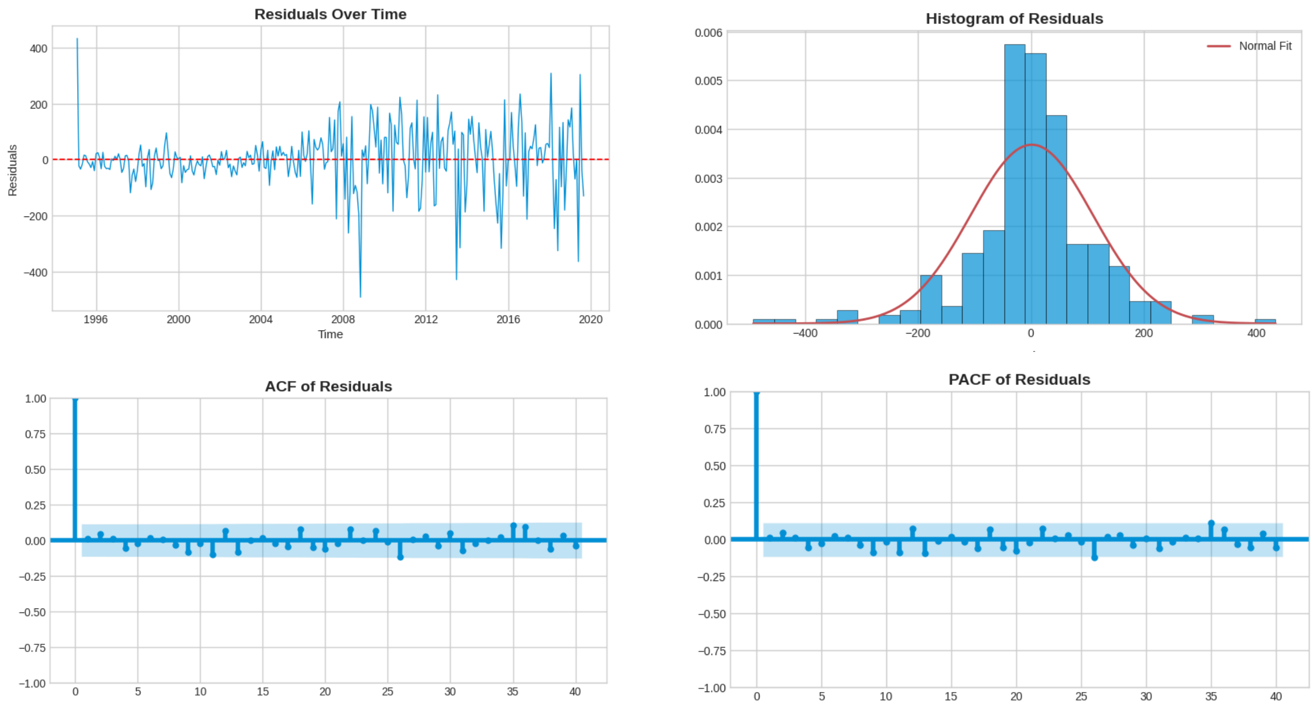
ACF results show that the autocorrelation value is very high at many lags and decreases slowly, which means that the IHSG data is highly non-stationary and has random walk characteristics, so differencing is needed to make it stationary.

PACF only shows a large spike at lag 1, while other lags are insignificant, indicating that after differencing, the data will follow an AR(1) pattern, so the indicated model is ARIMA with an AR component at lag 1.

The Augmented Dickey-Fuller (ADF) test is a statistical test to determine whether a time series is stationary or not.

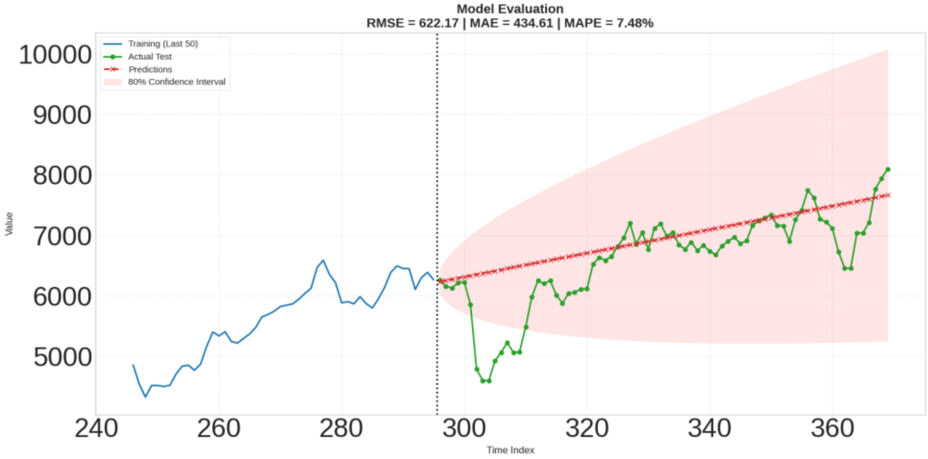
```
ADF Statistic: 0.442792
p-value: 0.983027
Critical Values:
        1%: -3.449
        5%: -2.870
        10%: -2.571
Failed to Reject Ho-Time Series is Non-Stationary
```

With the ADF test, it can be proven that the data is classified as **non-stationary**, so it can be concluded that the data has trends and seasonality



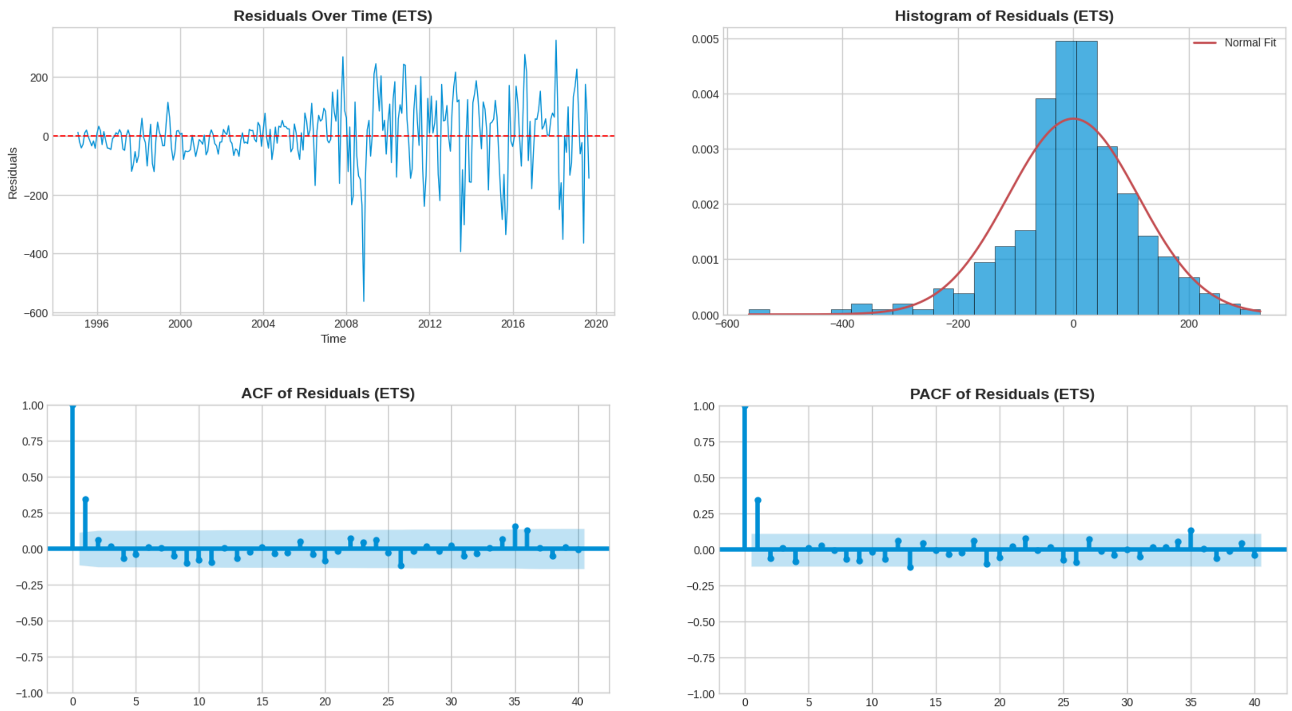
Model Diagnostics ARIMA Model

The residual ARIMA shows white noise characteristics: fluctuating around zero without a clear trend pattern, its distribution is close to normal, and there is no significant autocorrelation in the ACF and PACF (all spikes are within the 95% confidence interval). This indicates that the main data structure has been successfully captured by the model, so the model is considered adequate, even though there are indications of slightly changing volatility in certain periods.



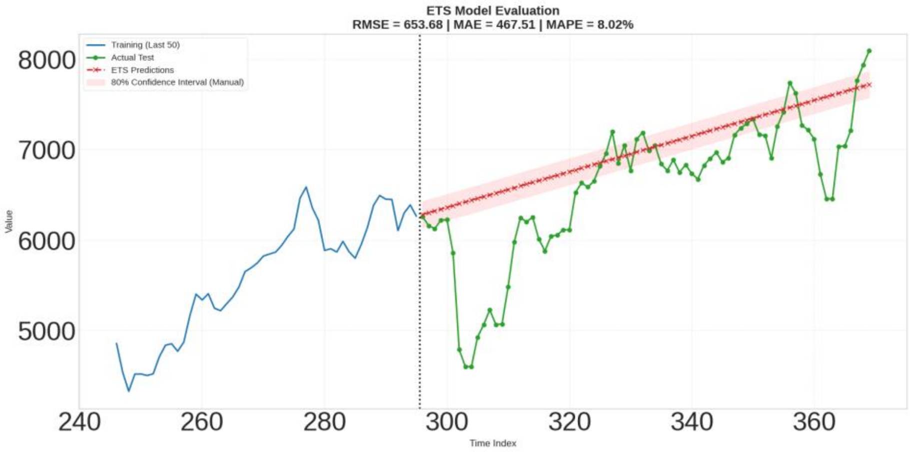
Model Evaluation ARIMA Model

Based on the evaluation model, ARIMA performance is quite good with RMSE = 622.17, MAE = 434.61, and MAPE = 7.48%, indicating a relative error rate below 10% (accurate category for economic/financial forecasting). The predictions follow the direction of the test data trend quite consistently, but there is a widening of the confidence interval as the horizon increases, indicating increased prediction uncertainty.



Model Diagnostics ETS Model

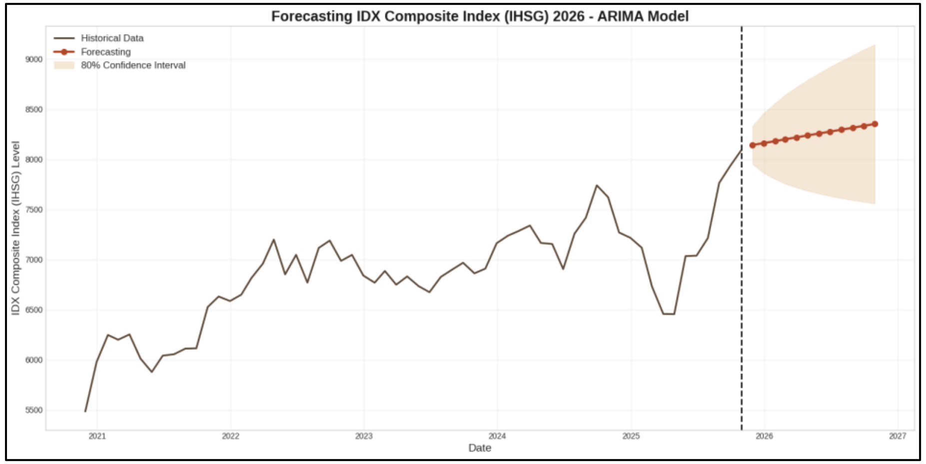
Based on ETS diagnostics, the residuals fluctuate around zero without any clear trend or seasonal patterns, indicating that the model has generally captured the main structure of the data. However, the residual distribution shows a slightly thick tail and is not quite normal (indicating outliers), as well as an increase in variance during certain periods, pointing to potential heteroscedasticity. In the ACF and PACF, several significant spikes still appear in the early lags, indicating residual autocorrelation that is not fully explained by the model.



Model Evaluation ETS Model

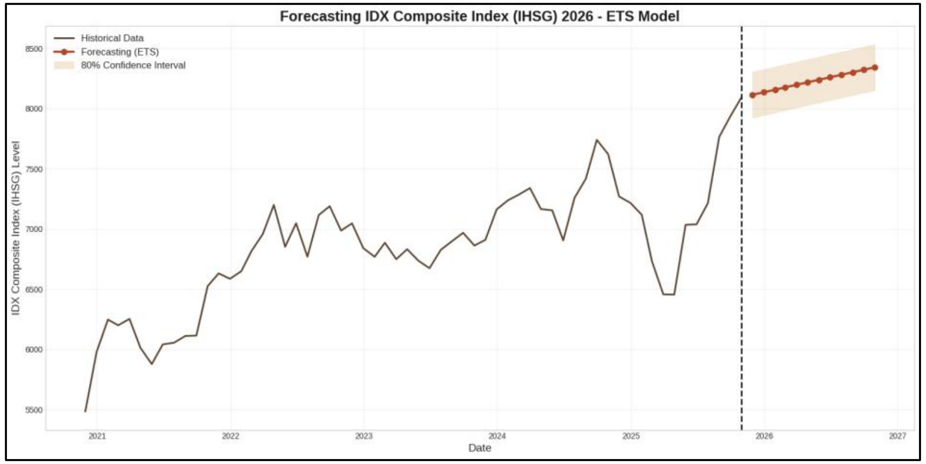
The ETS evaluation results show fairly good performance with RMSE = 653.68, MAE = 467.51, and MAPE = 8.02% (relative accuracy <10%), but numerically it is still slightly lower than ARIMA. Visually, ETS predictions follow the general trend of the test data, but tend to over-smooth, making them less responsive to sharp declines and short-term fluctuations, as seen in the large deviations at the beginning of the test period.

Forecasting IDX Composite Index (IHSG) 2026 ARIMA Model



Data Forecasting ARIMA Model			
Date	Forecast	Lower 80%	Upper 80%
11/30/2025	8142.227367	7952.507419	8331.947315
12/31/2025	8161.27287	7861.213808	8461.331931
1/31/2026	8180.318372	7800.742452	8559.894292
2/28/2026	8199.363875	7754.256446	8644.471303
3/31/2026	8218.409377	7716.45127	8720.367484
4/30/2026	8237.454879	7684.990976	8789.918783
5/31/2026	8256.500382	7658.229491	8854.771273
6/30/2026	8275.545884	7635.092494	8915.999274
7/31/2026	8294.591387	7615.243781	8973.939015
8/31/2026	8313.636889	7598.002424	9029.271355
9/30/2026	8332.682391	7583.451554	9081.913229
10/31/2026	8351.727894	7570.713699	9145.742089

Forecasting IDX Composite Index (IHSG) 2026 ETS Model



Data Forecasting ETS Model			
Date	Forecast	Lower 80%	Upper 80%
11/30/2025	8113.999	7923.391	8304.667
12/31/2025	8134.724	7944.116	8325.332
1/31/2026	8155.449	7964.84	8346.057
2/28/2026	8176.174	7985.566	8366.782
3/31/2026	8196.9	8006.291	8387.507
4/30/2026	8217.624	8027.016	8408.231
5/31/2026	8238.349	8047.74	8428.956
6/30/2026	8259.074	8068.465	8449.68
7/31/2026	8279.799	8089.19	8470.405
8/31/2026	8300.524	8109.915	8491.13
9/30/2026	8321.249	8130.64	8511.854
10/31/2026	8341.974	8151.366	8532.581

Discussion

To compare the two models, the most efficient method is to compare evaluation metrics, namely comparing the RMSE, MAE, and MAPE values between the ARIMA and ETS models. Lower RMSE, MAE, and MAPE values indicate better forecasting accuracy (Zhang, 2024).

RMSE	MAE	MAPE
622.1699	434.6055	7.48%

ARIMA Model Performance

RMSE	MAE	MAPE
653.6795	467.5080	8.02%

ETS Model Performance

Based on the metric results from both models, the ARIMA model value is better than the ETS model value by a small difference. This means the ARIMA model is better at predicting or forecasting.

Limitations

The Time Series method has limitations in performing analysis, including the following:

- It only looks at **Linear Relationships** and cannot see more complex relationships.
- It can only analyze one variable, known as **Univariate**, so other factors such as interest rates, GDP, exchange rates, and other macroeconomic factors cannot be included.

References

Zhang, H. (2024). Time-Series Analysis and Forecasting of S&P500 Index Based on ARIMA and ETS Model. *Highlights in Business, Economics and Management*, *24*, 306-316. https://doi.org/10.54097/anhgak96

More Information

If you want to see the **full project flow (including the code)**, you can visit **my GitHub account** or click this link.

https://github.com/mufadluqman/Time-Series-Analysis-and-Forecasting-of-IDX-Composite-Index-IHSG-Based-on-ARIMA-and-ETS-Model

https://github.com/mufadluqman