Forecasting CO2 Emission in Indonesia

Septa Dwi Cahya

https://www.linkedin.com/in/septadwicahya

Background

- Indonesia faces challenges in managing carbon dioxide (CO_2) emissions, especially from the industrial and transportation sectors. High CO_2 emissions can have significant impacts on global climate change as well as local air quality, with the potential to threaten human health and environmental sustainability.
- In this context, the forecasting model of CO_2 emissions in the future is important. The forecasting model can provide valuable information for stakeholders, researchers, and the public in planning mitigation strategies, optimizing resources, and estimating the impacts of policies to reduce CO_2 emissions.
- By understanding the historical trends of CO_2 emissions, we can develop the forecasting model to predict the level of CO_2 emissions in the future. It will be a contribution to the mitigation of climate change and the protection of air quality in the future.

Data Source

Emissions data are sourced from:

- https://data.world/makeovermonday/2019w22/workspace/file?filename=CO2+emissions+per+capita+per+country.csv
- https://data.worldbank.org/indicator/EN.ATM.CO2E.PC?end=2020&locations=ID&start=1990

The dataset will be used is CO2 emissions data in Indonesia from 1960 to 2020. The data will be tidied up first in excel. So, the next data cleaning process will not be too advanced.

Note: The unit for CO2 emissions is commonly measured in metric tons per year (t/yr)

Checking & Transforming Data

The information on the side shows that the Dataset contains 2 variables: 'year' (data type = integer) and 'CO_2' (data type = float).

The dataset also does not contain missing values and duplicate data.

So, we don’t need to do advanced data cleaning.

```
df.info()
✓ 0.0s

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 61 entries, 0 to 60
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   year    61 non-null     int64
 1   CO2     61 non-null     float64
dtypes: float64(1), int64(1)
memory usage: 1.1 KB
```

```
df.duplicated().any()
✓ 0.0s

False
```

Checking & Transforming Data

But, the variable 'year' (data type = integer) needs to be transformed to data type = datetime.

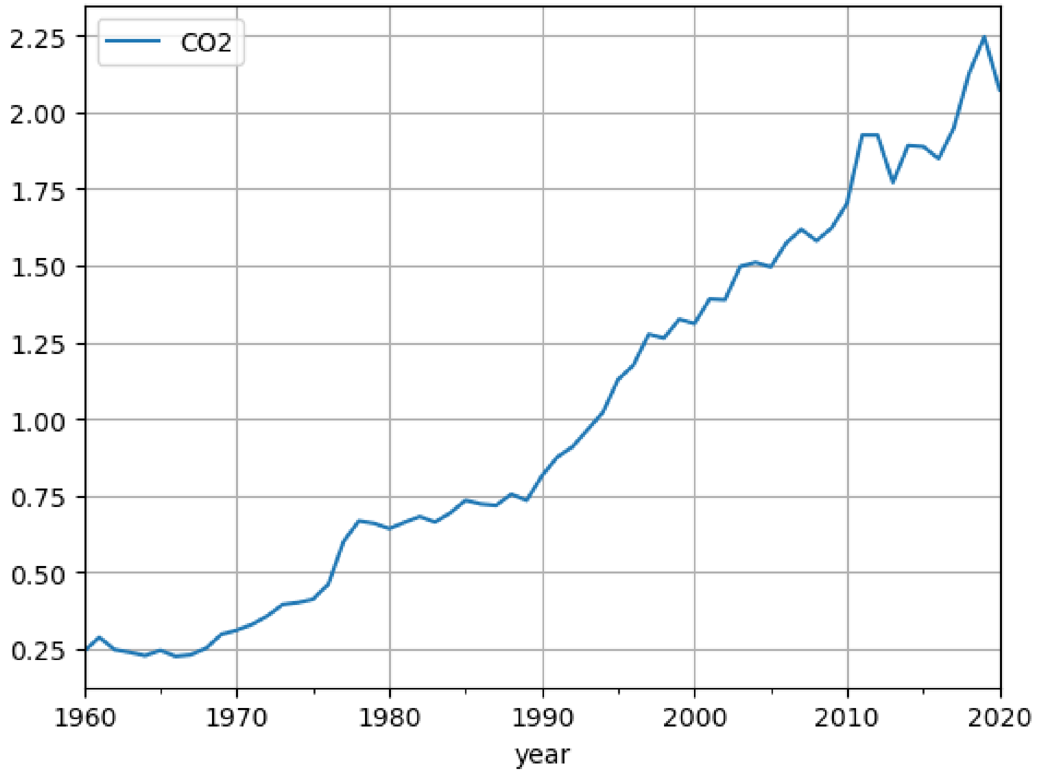
	year	CO2
0	1960	0.243805
1	1961	0.288720
2	1962	0.248447
3	1963	0.239681
4	1964	0.229361

	CO2
year	
1960-01-01	0.243805
1961-01-01	0.288720
1962-01-01	0.248447
1963-01-01	0.239681
1964-01-01	0.229361

Now, the Dataset can be proceed to the next analysis process.

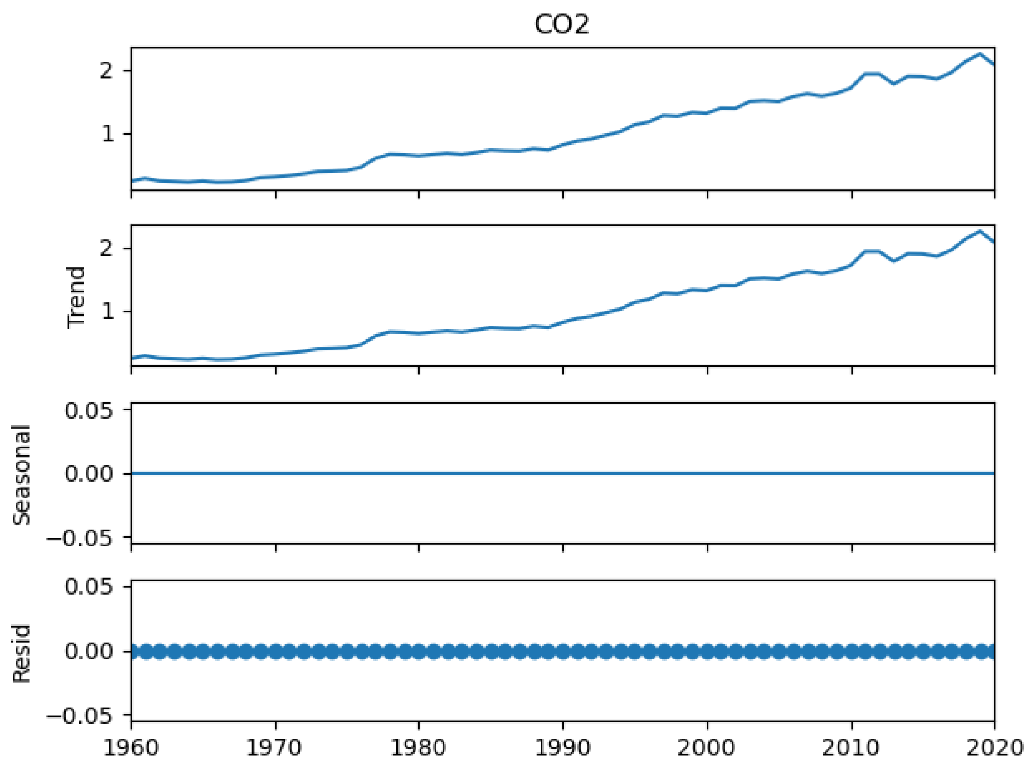
Data Exploration & Data Visualisation

CO2 emissions in Indonesia tended to increase from 1960 to 2020 (although in 2020 it decreased from the previous year).



Data Exploration & Data Visualisation

Seasonal Decomposition Plot



Trend: Overall increase in CO_2 emissions from year to year.

Seasonal: The seasonal graph is a straight line. It shows that there is no significant seasonal pattern in the data. So, CO_2 emissions do not show consistent periodic or cyclical changes over time.

Residual: The residual is around the zero line. It indicates that the long-term trends and seasonal patterns that have been identified are good enough to explain most of the variation in the data.

Data Exploration & Data Visualisation

Augmented Dickey-Fuller (ADF) Test

```
The Result of testing Dickey-Fuller
--------------------------------
Test Statistic                 -2.534039
p-value                         0.107410
Lags Used                       8.000000
Number of Observation Used     51.000000
Critical Value (1%)            -3.565624
Critical Value (5%)            -2.920142
Critical Value (10%)           -2.598015
dtype: float64
```

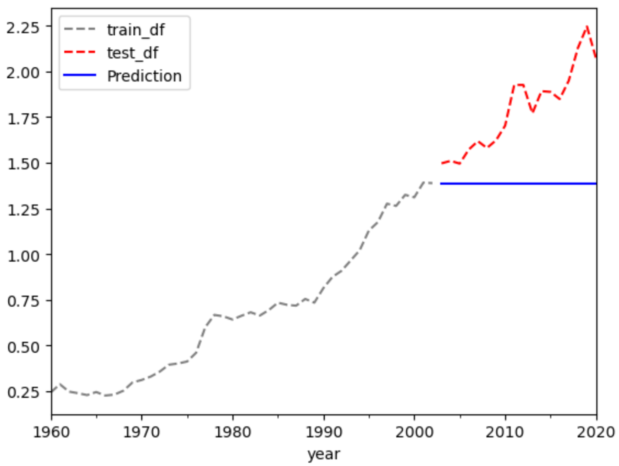
- Since the *p-value* (0.107410) is greater than the significance level (e.g. 0.05), we fail to reject the null hypothesis.
- The null hypothesis in the ADF test is that the data is non-stationary.
- So, the conclusion from this result is that there is not enough evidence to state that the time series data is stationary.

Time Series methods can be used for non-stationary data: ARIMA (Autoregressive Integrated Moving Average), Exponential Smoothing Models (including Single Exponential Smoothing, Double Exponential Smoothing), etc.

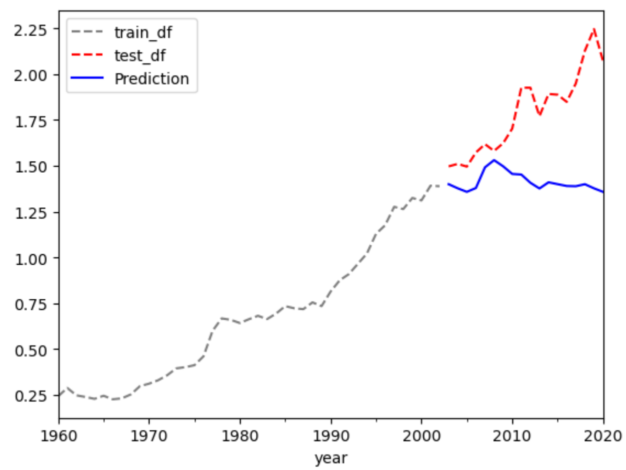
Modeling & Analysing Data

Modelling the dataset will use Single Exponential Smoothing, Double Exponential Smoothing, and ARIMA methods. Before modelling, the dataset will be divided into 2: Training Data (70%) and Testing Data (30%).

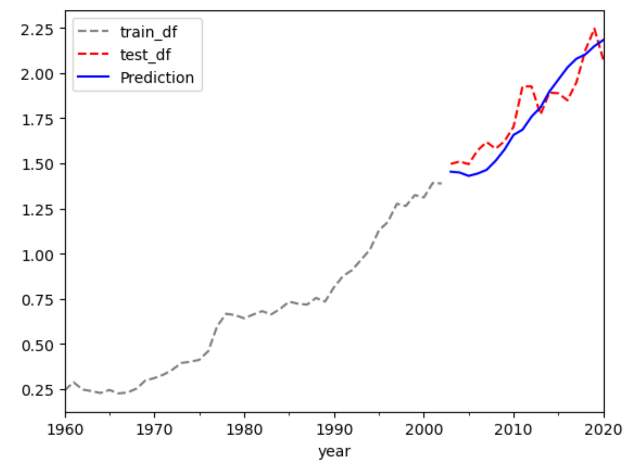
Single Exponential Smoothing



Double Exponential Smoothing



ARIMA

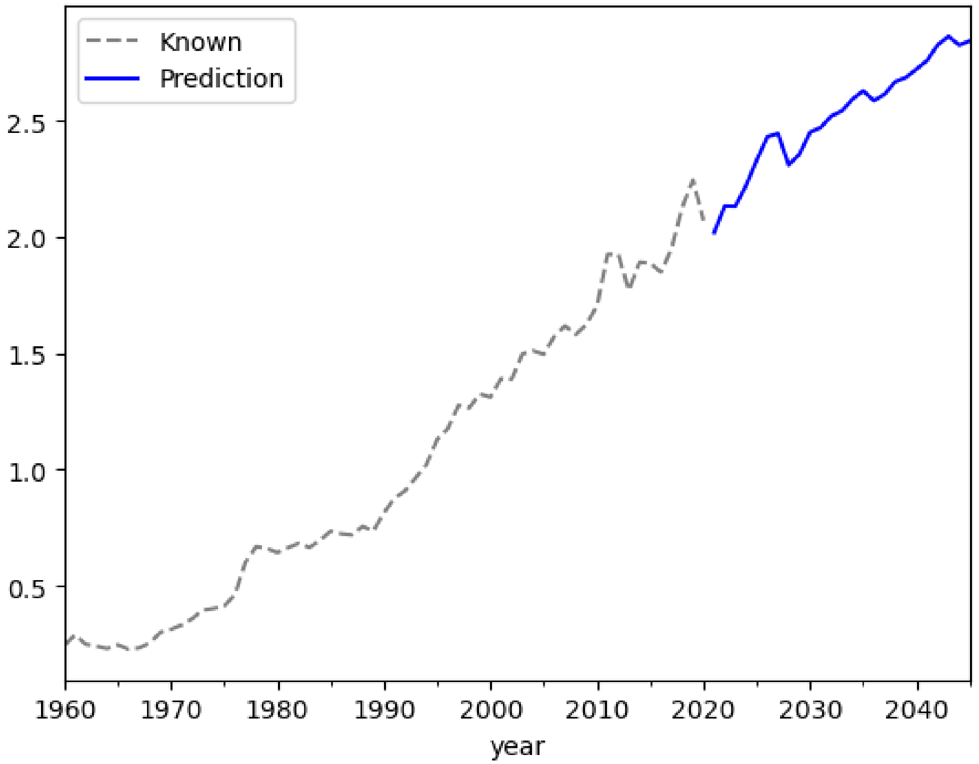


Based on the graph above and the RMSE & MAPE values, it can be concluded that the best model for forecasting CO2 emissions in Indonesia is ARIMA.

	RMSE	MAPE
Model		
ARIMA	0.111473	0.052077
Double Exp Smoothing	0.448010	0.197776
Single Exp Smoothing	0.460104	0.212895

Forecasting

Forecasting Indonesia’s CO2 emissions in 2045



For more Forecasting in other years, you can try in this link (It provide in dynamic graph)

https://capstone-project-forecast-co2-indonesia-septadwicahya.streamlit.app/

Insight & Recommendation

- Based on the forecasting model, CO_2 emissions in Indonesia tend to increase from year to year. This indicates that there is continued growth in emission levels. So, it's crucial to address this issue to mitigate the potential negative impacts on climate change.
- Here are some strategies to tackle the increase in CO_2 emissions in Indonesia:
  - Development of Renewable Energy
  - Enhance Energy Efficiency
  - Public Transportation and Sustainable Mobility
  - Strengthen Environmental Regulations
  - Encourage Afforestation and Reforestation