Forecasting CO_2 Emissions for G20 Countries Using Time Series and Deep Learning Models

Name: Satyanadh Reddy
SJSU ID: 016037957

1. Introduction:

This project mainly focuses on predicting global C02 emissions. Global CO2 emissions have become the main reason for the climatic changes all over the world. These emissions can cause serious impacts on climatic changes, air pollution and many health related issues. The countries present in the G20(Group of 20 countries) meet regularly to make decisions and make policies on trade, climate change and health of the people. Among G20 countries the United States, India, and China are among the highest contributors to CO2 emissions globally. Accurate forecasting of the CO2 emissions is needed for developing effective environmental policies and mitigation strategies.

2. Objective:

This study aims to develop effective predictive models for CO2 emissions in G20 countries using time series forecasting techniques. We focus on evaluating the various Machine Learning, Deep Learning and statistical models to identify the best performing model for predicting the CO2 emissions. The main objective of this study is to develop a time series forecasting models for G20 countries CO2 emissions from 1980 to 2019 using univariate time series data**.** Evaluate the models based on some performance metrics, including MSE, RMSE, MAE, and others.

3. Problem Statement:

Global CO2 emissions have been steadily increasing, endangering both human health and the environment. It is essential to project these emissions for the ensuing ten years in order to create policies and manage the environment. Forecasting G20 Countries CO2 emissions using univariate time series data from 1980 to 2020 is the problem this study attempts to solve. Various forecasting models, such as SARIMAX, Holt-Winters, linear regression, random forest, and LSTM, are evaluated for their performance.

4. Methodology:

4.1 Dataset:

The dataset includes a number of G20 members, including Canada, Australia, the United Kingdom, Germany, Japan, and others, along with numerous significant CO2 emitters, including China, India, the USA, and Brazil. It offers a thorough picture of CO2 emissions worldwide, facilitating cross-national comparisons and analysis of temporal trends. The data is univariate in nature and illustrates how each nation's CO2 emissions have changed over time. The collection is comprehensive and contains statistics on the several sectors within a country that contribute to emissions. The dimensions of the dataset is 120 rows x 37 columns.

Columns:

ISO: The ISO codes for every nation are listed in this column.
Country: The names of each nation are listed in this column.
Data Source: The data source for the emissions data is shown in this column.
Sector: The industrial or economic sector responsible for the emissions is indicated in this column.
Gas: The type of gas being measured is indicated in this column, and in this dataset, that gas is CO2.
Unit: The metric ton is the unit of measurement that is specified in this column.
Yearly Columns: The total annual CO2 emissions for each year, expressed in metric tons, are shown in each column from 1990 to 2020.

4.2 Descriptive Statistics:

	ISO	Country	Data source	Sector	Gas	Unit	2020	2019	2018	2017	...	1999	1998	1997	1996	1995	1994
0	CHN	China	GCP	Total fossil fuels and cement	CO2	$MtCO_2e$	10914.01	10721.04	10353.93	10011.15	...	3557.27	3364.59	3515.59	3508.82	3361.64	3103.74
1	CHN	China	GCP	Coal	CO2	$MtCO_2e$	7679.55	7523.17	7316.40	7163.32	...	2580.99	2444.77	2609.56	2660.40	2569.75	2380.59
2	USA	United States	GCP	Total fossil fuels and cement	CO2	$MtCO_2e$	4714.63	5262.15	5377.80	5212.16	...	5803.92	5733.28	5688.14	5612.98	5425.26	5364.28
3	EUU	European Union (27)	GCP	Total fossil fuels and cement	CO2	$MtCO_2e$	2629.97	2904.16	3046.24	3119.78	...	3590.36	3645.83	3655.53	3721.91	3637.06	3590.76
4	IND	India	GCP	Total fossil fuels and cement	CO2	$MtCO_2e$	2421.55	2612.89	2593.06	2426.61	...	950.46	875.77	858.01	823.62	760.46	714.06

Fig-1.1 Head of the dataset.

```
#   Column        Non-Null Count  Dtype
--  ------        --------------  -----
0   ISO           120 non-null    object
1   Country       120 non-null    object
2   Data source   120 non-null    object
3   Sector        120 non-null    object
4   Gas           120 non-null    object
5   Unit          120 non-null    object
6   2020          120 non-null    float64
7   2019          120 non-null    float64
8   2018          120 non-null    float64
9   2017          120 non-null    float64
10  2016          120 non-null    float64
11  2015          120 non-null    float64
12  2014          120 non-null    float64
13  2013          120 non-null    float64
14  2012          120 non-null    float64
15  2011          120 non-null    float64
16  2010          120 non-null    float64
17  2009          120 non-null    float64
18  2008          120 non-null    float64
19  2007          120 non-null    float64
20  2006          120 non-null    float64
21  2005          120 non-null    float64
22  2004          120 non-null    float64
23  2003          120 non-null    float64
24  2002          120 non-null    float64
25  2001          120 non-null    float64
26  2000          120 non-null    float64
27  1999          120 non-null    float64
28  1998          120 non-null    float64
29  1997          120 non-null    float64
30  1996          120 non-null    float64
31  1995          120 non-null    float64
32  1994          120 non-null    float64
33  1993          120 non-null    float64
34  1992          120 non-null    float64
35  1991          120 non-null    float64
36  1990          120 non-null    float64
```

Fig-1.2 Info of the dataset.

4.3 Data Visualization:

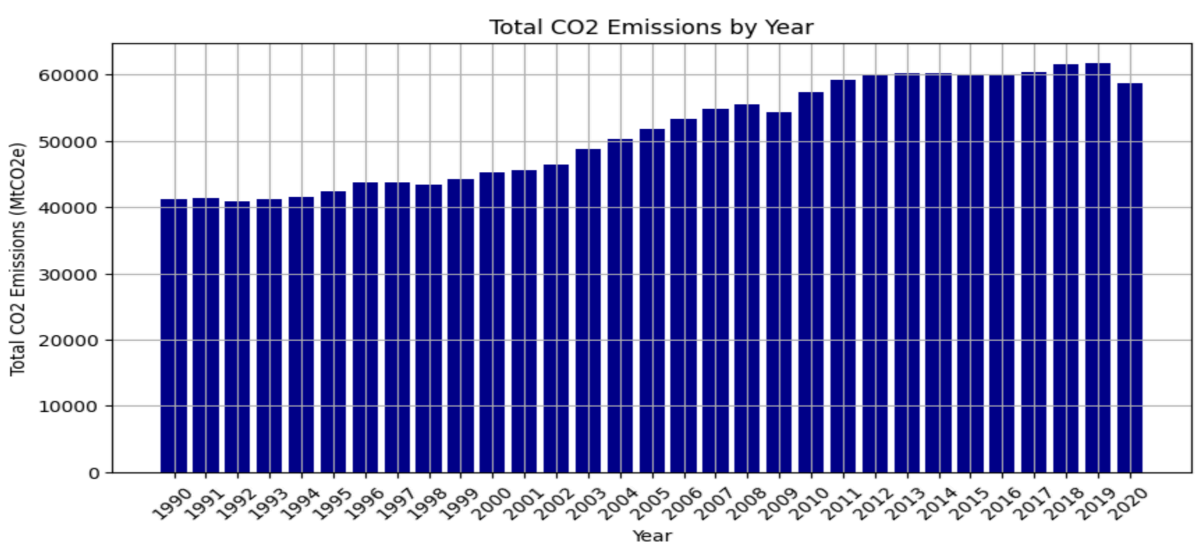


Fig-1.3 Total CO2 Emissions by G20 Countries by each Year

From the above figure we can observe there was a significant increase in the CO2 emissions from 1990 till 2019 by all countries.

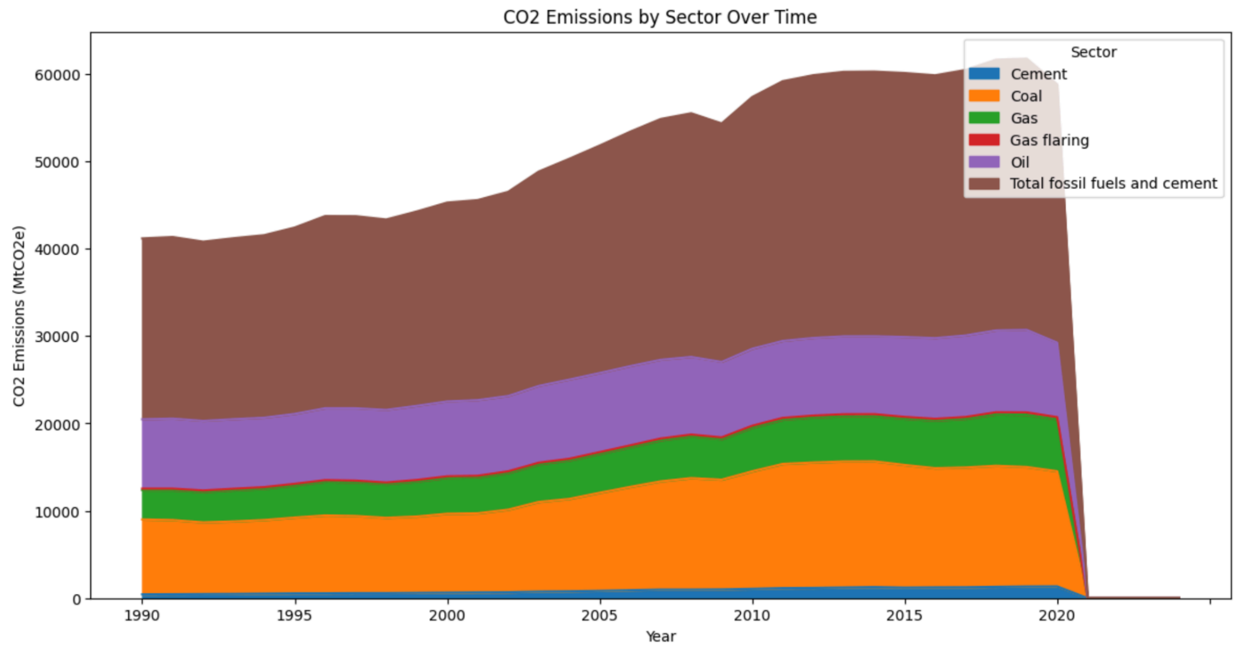


Fig-1.4 Total CO2 Emissions by each sector from 1990 to 2020.

From the above figure we can observe that among all Total fossil fuels and cement contributes more for CO2 emissions and is followed by oil and Gas flaring. Cement is the least among all.

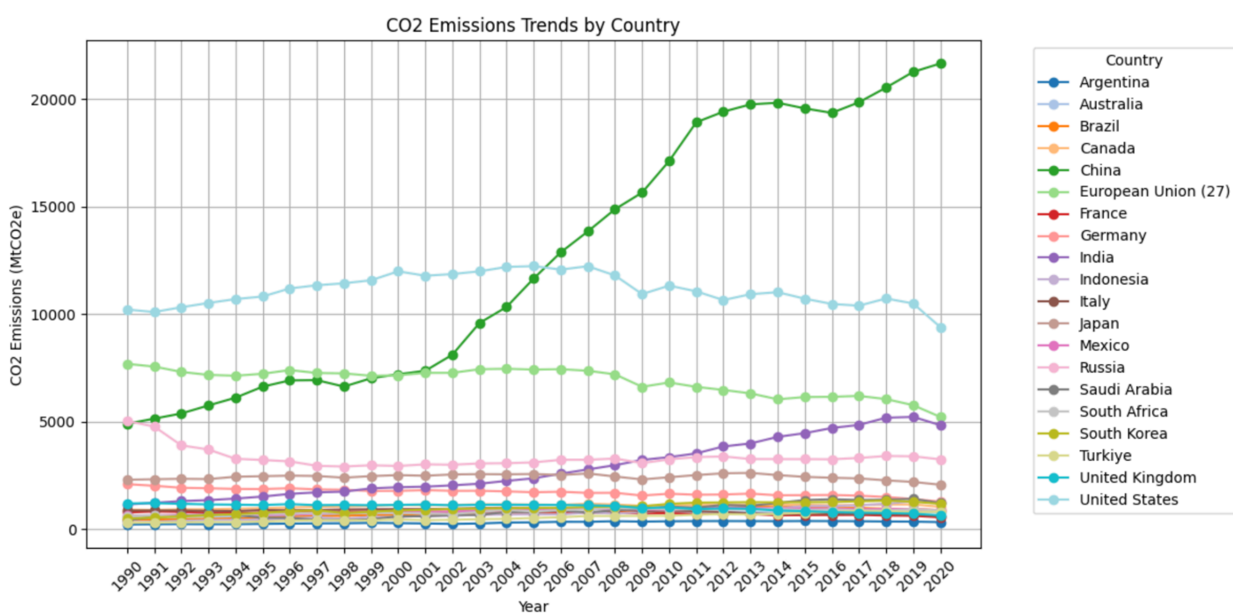


Fig-1.5 Total CO2 Emissions by each country from 1990 to 2020.

From the above figure we can observe which country is emitting more and which other countries are emitting less. Among all, China is emitting more and is followed by the United states.

4.4 Data Pre-processing:

- **Handling Null/Missing Values:**

After reviewing the data from 1980 to 2020, I found out that there aren't any missing or empty values. This means there is no need to deal with filling in any nan/null values in the dataset. This clean dataset gives us a solid foundation to work from, without having to worry about filling in any null/nan values.

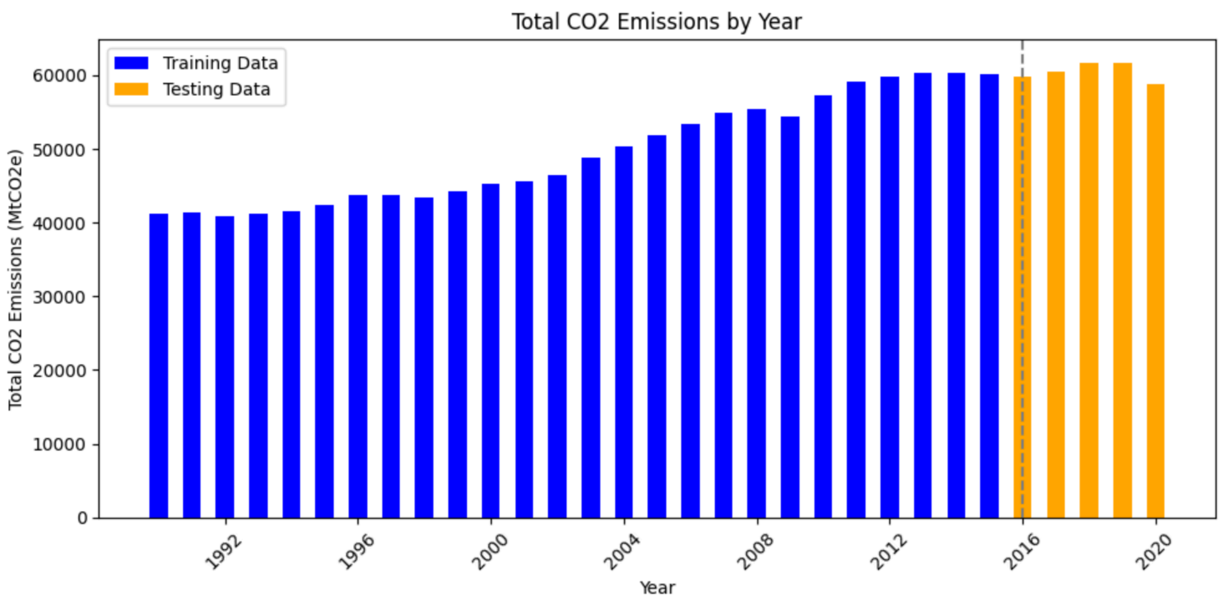
- **Data Preparation for Training & Testing :**

1. I focused on the CO_2 emissions data from 1980 to 2020.
2. Columns Dropped: I removed all columns except for the emission values, including:
    - **ISO**
    - **Data Source**
    - **Sector**
    - **Gas**
    - **Unit**
3. I summed up the emissions across all sectors within each country.
4. The final dataset contained:
    - Total CO_2 emissions for each country
    - Country names as identifiers.

Country Year	Argentina	Australia	Brazil	Canada	China	European Union (27)	France	Germany	India	Indonesia	Italy	Japan	Mexico	Russia
1990-01-01	224.16	554.00	433.62	913.87	4905.63	7691.51	781.86	2098.55	1156.00	310.16	872.89	2303.13	634.08	5034.72
1991-01-01	235.01	556.76	455.04	897.05	5144.14	7565.76	830.67	2023.67	1230.72	349.73	871.73	2326.88	660.26	4778.03
1992-01-01	242.46	566.62	463.25	925.04	5383.99	7316.93	811.03	1929.05	1310.90	398.84	872.44	2347.33	665.56	3901.94
1993-01-01	235.55	575.10	484.37	926.15	5757.45	7179.42	772.33	1908.86	1354.60	430.21	857.29	2333.61	676.14	3709.96
1994-01-01	236.94	584.32	503.91	954.81	6117.73	7139.88	758.42	1875.84	1428.13	438.74	845.99	2443.88	703.69	3277.22
1995-01-01	250.49	607.13	533.25	980.43	6630.43	7230.76	768.02	1869.20	1520.91	444.81	893.87	2468.58	662.28	3221.77
1996-01-01	262.26	620.77	574.37	1013.17	6922.74	7401.76	801.00	1909.26	1647.25	506.14	882.68	2493.68	683.53	3158.27
1997-01-01	273.05	637.35	609.55	1041.69	6935.56	7268.02	786.03	1852.80	1716.02	559.88	894.19	2479.06	725.36	2950.54

4.5 Data Splitting:

To prepare the data for analysis, I divided the CO_2 emissions dataset into training and testing sets. I allocated 36 years (from 1980 to 2015) for training and used the remaining 5 years (from 2016 to 2020) for testing. This corresponds to approximately 87.5% of the data for training and 12.5% for testing.



Training and Testing data Split

4.6 Models:

- **Auto-Regressive Integrated Moving Average (ARIMA):**

This model is a popular statistical method used for time series forecasting. This model has three components mainly:

> **Auto-Regressive (AR):** Incorporates the influence of past values (lags).
> **Integrated (I):** Involves differencing the data to remove trends.
> **Moving Average (MA):** Models the error as a linear combination of past error terms.

Model Parameters :

In this project, We manually set the parameters of the ARIMA model as follows:

p: The number of lag observations included in the model, which represents the autoregressive terms. We set “P = 1”.
d: The number of times that the raw observations are differenced to make the time series stationary. We set “d = 1”.
q: The size of the moving average window, which represents the moving average terms. We set “q = 1”.

This model assumes that the data must be stationary, which means that the series statistical characteristics, such as mean, variance, and autocorrelation, do not vary over time. This is one of the fundamental presumptions of ARIMA. If the data is not stationary before fitting the model we should mention it by giving the “d = 1”.

As the ARIMA model depends on recurring patterns in the data to produce precise forecasts, stationarity is essential. The model may produce inaccurate or misleading predictions if the data is not steady.

● Holt-Winters Model

An expansion of the Exponential Smoothing technique that takes trend and seasonality in time series data into consideration is the Holt-Winters model, sometimes referred to as Triple Exponential Smoothing. When forecasting time series data with seasonal fluctuations and trends, this model is especially helpful.

- **Model Components**

Level: The baseline value of the series, which adjusts over time.
Trend: The increasing or decreasing value in the series, which adjusts over time.
Seasonality: The repeating pattern or cycle in the series, which adjusts over time.

In this project, I’m using the Holt-Winters model with the following settings:

Seasonal: We used an additive seasonal component (**seasonal='add'**), suitable for series with constant seasonal variations.
Seasonal Periods: We set **seasonal_periods=12** to capture yearly seasonality.
Trend: We used an additive trend component (**trend='add'**), suitable for series with linear trends.
Damped Trend: We set **damped_trend=False** to allow the trend to continue indefinitely.

- ## Linear Regression Model

A basic statistical method for simulating the relationship between a dependent variable and one or more independent variables is called linear regression. It assumes a linear relationship between the variables, which makes it a simple yet effective tool for forecasting and trend analysis, especially when the relationship is straightforward and linear.

Model Components :

Dependent Variable (y): The variable we want to predict or explain (in this case, CO_2 emissions).
Independent Variable (X): The variable(s) used to make predictions (in this case, the year).

Limitations :

Linear Assumption: It assumes a linear relationship, which might not capture more complex patterns or interactions.
Outlier Sensitivity: The model is sensitive to outliers, which can skew the results and impact accuracy.
Overfitting: With too many independent variables, the model can overfit the training data, reducing generalization to new data.

- ## Random Forest Model

An ensemble learning method called Random Forest is applied to tasks involving regression and classification. During training, it builds several decision trees, and for regression problems, it outputs the mean of the predictions. Random Forest is a suitable option for CO_2 emissions forecasting because it is robust, controls both linear and nonlinear relationships well, and can handle enormous datasets.

Model Components :

In this project, I configured the Random Forest model with the following parameters:

Number of Trees (n_estimators):

- This parameter specifies the number of decision trees in the forest.

- We set **n_estimators=100** to create a forest with 100 trees.
- Having more trees generally improves the model's performance but also increases computational cost.

Max Depth (max_depth):

- This parameter determines the maximum depth of each tree.
- If not specified, trees grow until all leaves are pure or until they contain fewer than **min_samples_split** samples.
- We did not explicitly set this parameter, allowing the trees to grow as needed.

Min Samples Split (min_samples_split):

- This parameter defines the minimum number of samples required to split an internal node.
- We did not explicitly set this parameter, so it defaults to 2.
- A higher value prevents the model from overfitting by making splits only when there is sufficient data.

Min Samples Leaf (min_samples_leaf):

- This parameter specifies the minimum number of samples required to be at a leaf node.
- We did not explicitly set this parameter, so it defaults to 1.
- A higher value makes the model more conservative by preventing overly specific splits.

● Long Short-Term Memory (LSTM) Model

Recurrent neural networks (RNNs) with Long Short-Term Memory (LSTM) are particularly good at learning and predicting time series data. Because LSTM networks can sustain and refresh a "memory" of past inputs over extended periods of time, they are well-suited for sequential data, as they do not encounter the vanishing gradient issue that typical RNNs do.

Model Components :

Sequential Model:

- I used a Sequential model to stack the layers of our LSTM network sequentially.

LSTM Layer:

- The LSTM layer has 50 memory units, which are the number of neurons in the layer.

- The input shape for the LSTM layer is **(look_back, 1)**, indicating the window size (number of time steps) and one feature.

Dense Layer:

- The Dense layer has one neuron, corresponding to the output for the next time step.

Compilation:

- We compiled the model using **mean_squared_error** as the loss function and **adam** as the optimizer.

Evaluation Metrics:

To evaluate the performance and accuracy of the different models we used, we used a number of evaluation indicators in our analysis. Every indicator offers an alternative viewpoint on the prDescription:

Mean Squared Error (MSE)

- The Mean Squared Error (MSE) measures the average of the squares of the errors, which are the differences between the actual and predicted values.
- It squares the error to give more weight to larger differences.
- A lower MSE indicates better model performance.

Root Mean Squared Error (RMSE)

- The Root Mean Squared Error (RMSE) is the square root of the MSE.
- It provides a measure of the average magnitude of the error.
- RMSE is often used because it shares the same units as the original data, making it more interpretable.
- Like MSE, a lower RMSE indicates better model performance.

Mean Squared Logarithmic Error (MSLE)

- The Mean Squared Logarithmic Error (MSLE) measures the ratio between the predicted and actual values, taking the logarithm of both.
- It is useful when dealing with exponential growth data.
- MSLE penalizes underestimates more than overestimates, and smaller MSLE values indicate better performance.

Mean Absolute Error (MAE)

- The Mean Absolute Error (MAE) measures the average magnitude of the errors without considering their direction.
- It is the average absolute difference between the predicted and actual values.
- MAE is robust to outliers, and smaller MAE values indicate better performance.

Median Absolute Error (MedAE)

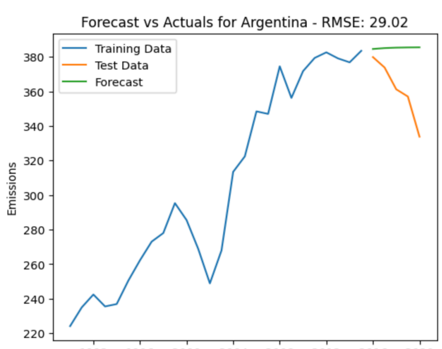
- The Median Absolute Error (MedAE) is similar to MAE, but it uses the median instead of the mean.
- It is particularly useful for datasets with skewed error distributions or outliers, as it gives a robust measure of central tendency.
- A lower MedAE indicates better model performance.

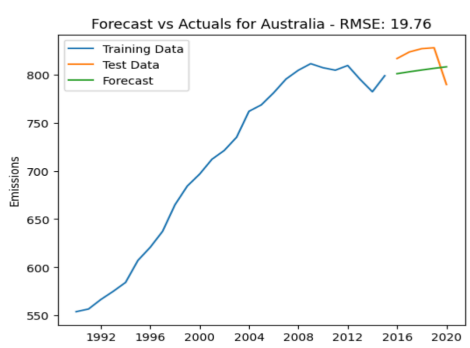
5. Results:

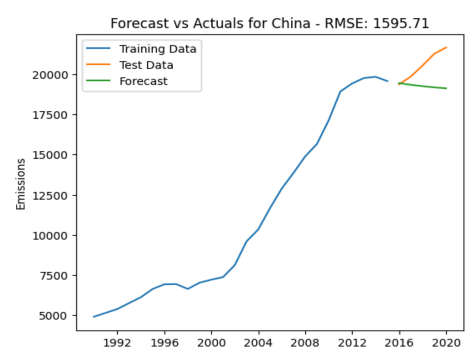
In this project I predicted the CO_2 emissions on the test data for each country individually so that they will not be affected by trends from one country to another country.

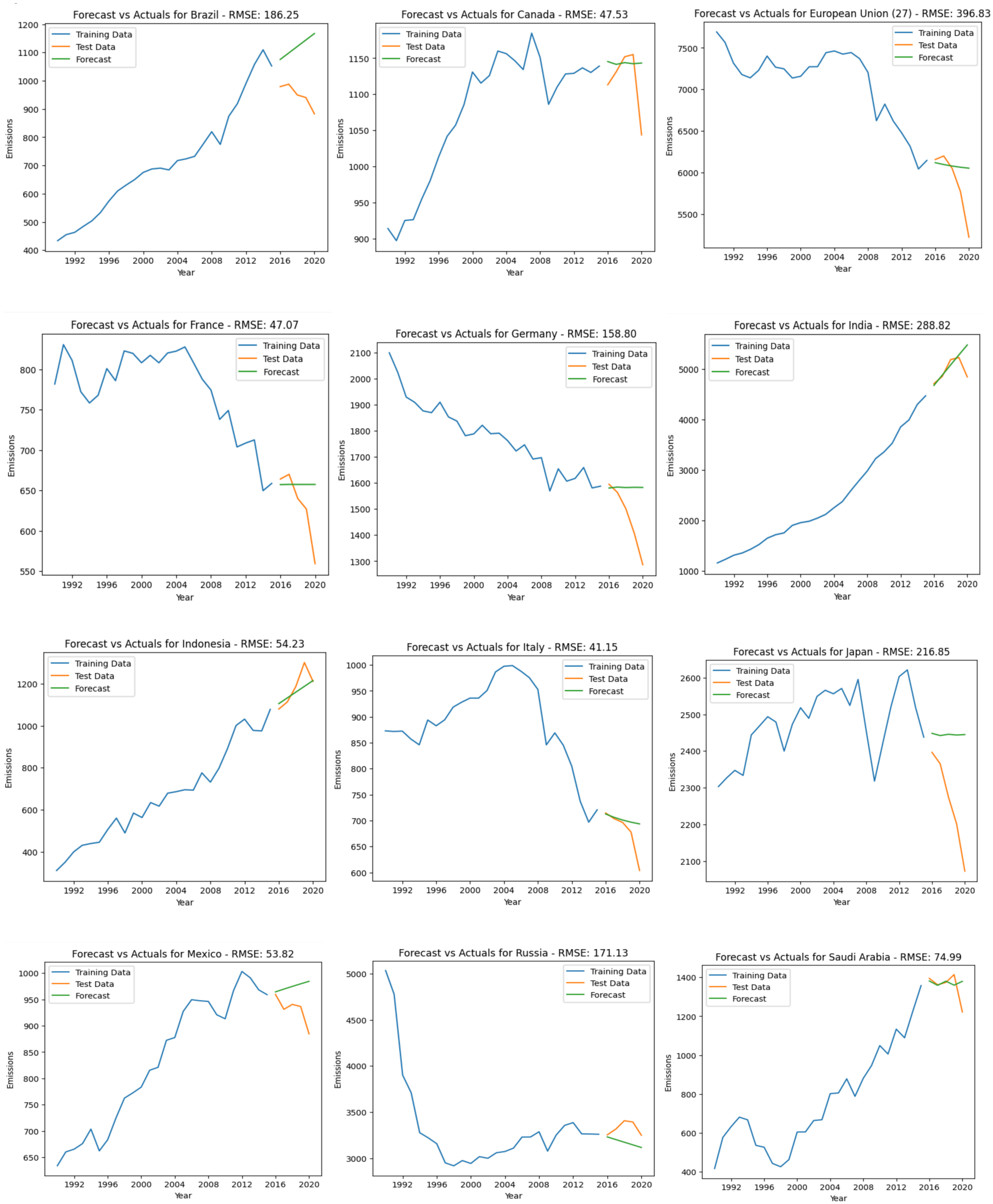
Arima Model -

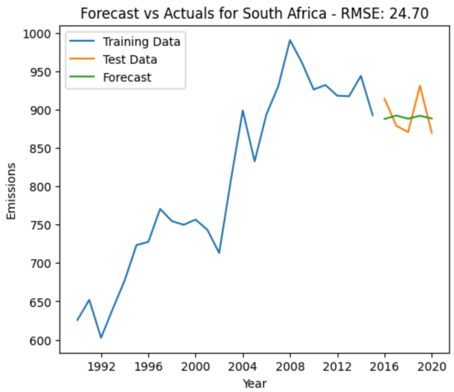
Below are the plots which shows the actual vs predicted for each country

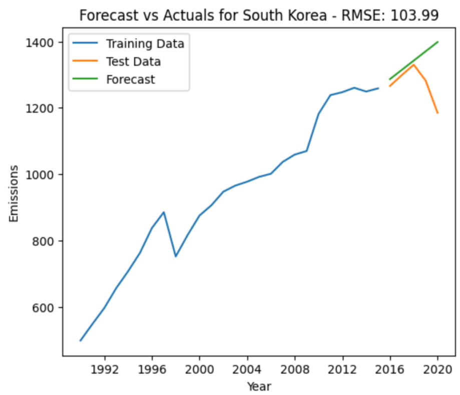


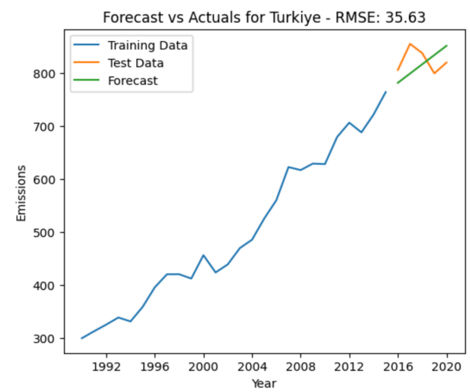


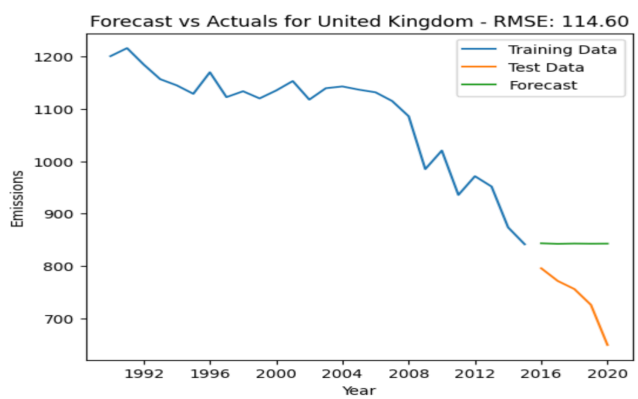



Forecast vs Actuals for Brazil - RMSE: 186.25
Forecast vs Actuals for Canada - RMSE: 47.53
Forecast vs Actuals for European Union (27) - RMSE: 396.83
Forecast vs Actuals for France - RMSE: 47.07
Forecast vs Actuals for Germany - RMSE: 158.80
Forecast vs Actuals for India - RMSE: 288.82
Forecast vs Actuals for Indonesia - RMSE: 54.23
Forecast vs Actuals for Italy - RMSE: 41.15
Forecast vs Actuals for Japan - RMSE: 216.85
Forecast vs Actuals for Mexico - RMSE: 53.82
Forecast vs Actuals for Russia - RMSE: 171.13
Forecast vs Actuals for Saudi Arabia - RMSE: 74.99
Training Data
Test Data
Forecast
Emissions
Year









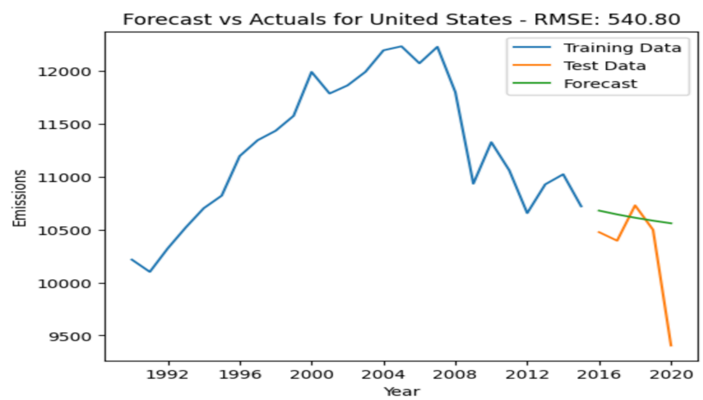


Table having different evaluation metrics for each country:

Country	MSE	RMSE	MSLE	MAE	MedAE
Argentina	842	29	0.01	24	24
Australia	391	20	0.00	20	21
Brazil	34690	186	0.03	173	172
Canada	2259	48	0.00	33	13
China	2546281	1596	0.01	1302	1288
European Union	157476	397	0.00	258	103
France	2215	47	0.01	33	17
Germany	25216	159	0.01	118	81

India	83417	289	0.00	171	54
Indonesia	2940	54	0.00	38	26
Italy	1693	41	0.00	24	5
Japan	47022	217	0.01	183	171
Mexico	2896	54	0.00	44	38
Russia	29285	171	0.00	150	133
SaudiArabia	5623	75	0.00	47	13
South Africa	610	25	0.00	23	19
South Korea	10814	104	0.01	70	21
Turkey	1269	36	0.00	33	32
United Kingdom	13133	115	0.02	103	87
United States	292462	541	0.00	363	204

The error metrics for the ARIMA model are displayed in the above table for each of the G20 nations. The findings show significant differences in the prediction errors of CO_2 emissions. Larger errors are seen in nations like China and the US, because of their intricate emission patterns and more developed economies, but lesser errors are seen in nations like Italy and Argentina, suggesting more stable or predictable emission trends. The evaluation metrics—MSE, RMSE, MSLE, MAE, and MedAE, among others—provide a thorough overview of the model's performance across a range of scales and error kinds, enabling us to efficiently evaluate and contrast the forecast's accuracy.

Holt winter Model -

Below plot shows the actual values vs predicted values in the dotted line for all countries:

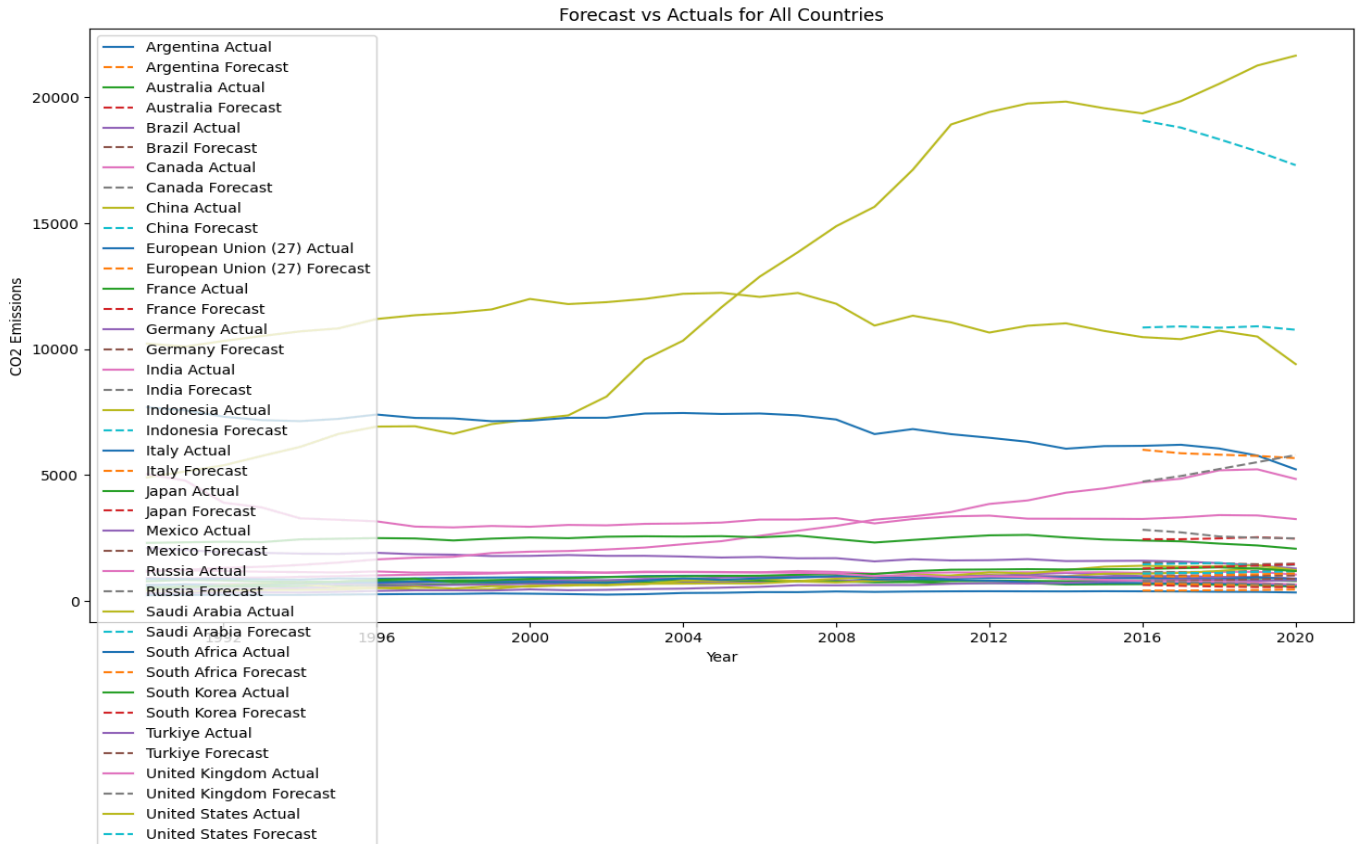


Table having different evaluation metrics for each country:

Country	MSE	RMSE	MSLE	MAE	MedAE
Argentina	5080	71	0.03	64	63
Australia	805	28	0.00	17	6
Brazil	35274	187	0.03	172	156
Canada	5069	71	0.00	48	38
China	7314398	2704	0.02	2258	2198
European Union	79153	281	0.00	239	246

France	3446	58	0.01	52	69
Germany	10054	100	0.01	80	71
India	198485	445	0.01	283	106
Indonesia	6050	77	0.00	57	48
Italy	1001	31	0.00	27	30
Japan	63997	252	0.01	214	206
Mexico	5824	76	0.01	66	63
Russia	525049	724	0.06	704	765
SaudiArabia	16371	127	0.01	108	118
South Africa	19226	138	0.02	128	125
South Korea	17282	131	0.01	92	26
Turkey	2249	47	0.00	43	50
United Kingdom	964	31	0.00	23	17
United States	488523	698	0.00	555	407

There are significant differences in forecasting accuracy among the G20 countries according to the Holt Winter model's error measures. While countries like the United Kingdom and Italy show lesser mistakes, indicating more predictable trends, China and the United States exhibit bigger errors, indicating more complex emission patterns. According to the measures, countries with a wider range of economic activity and greater sizes typically have higher forecast errors, which is indicative of the complexity of their emissions. On the other hand, countries that are smaller or more stable show less errors.

Linear Regression:

Below plot shows the actual values vs predicted values in the dotted line for all countries:

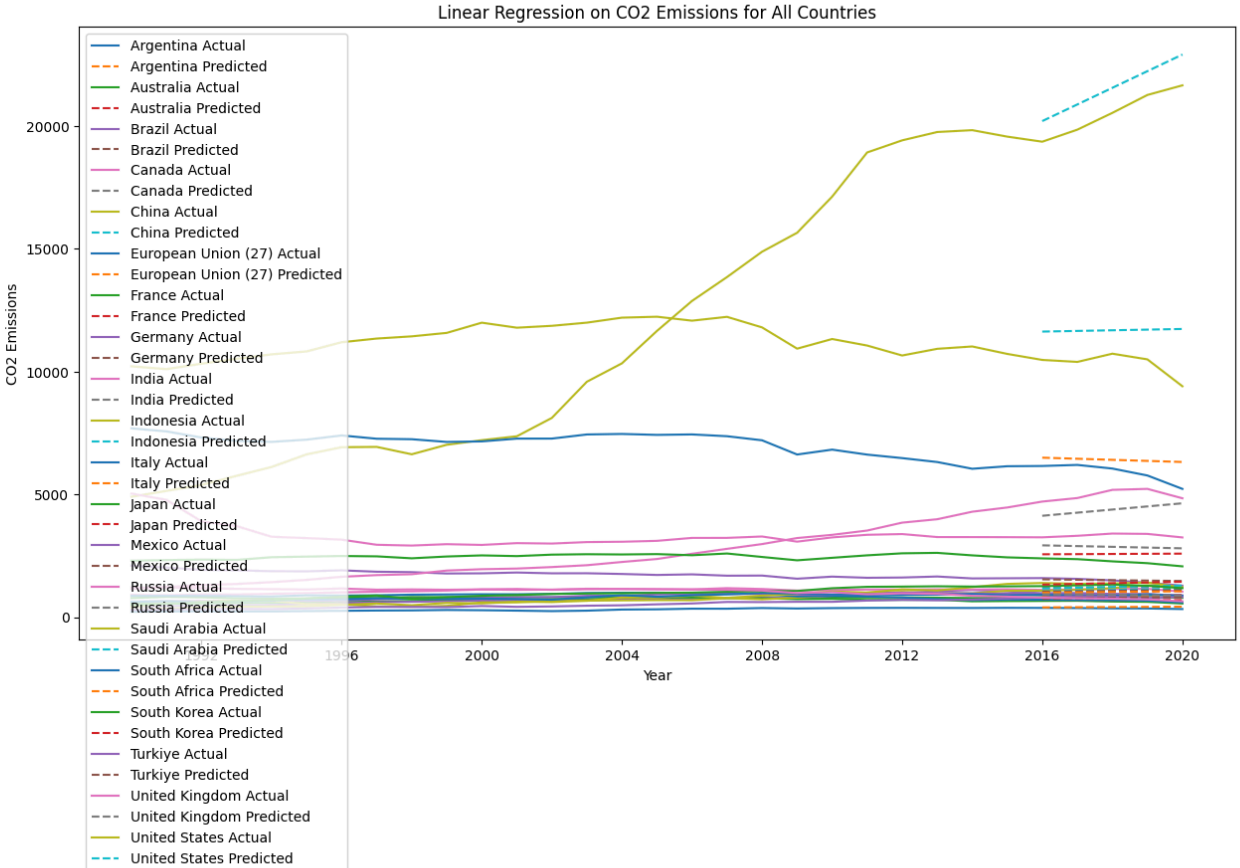


Table having different evaluation metrics for each country:

Country	MSE	RMSE	MSLE	MAE	MedAE
Argentina	3639	60	0.02	54	54
Australia	5917	76	0.01	72	62
Brazil	27266	165	0.03	149	147
Canada	13296	115	0.01	104	83
China	1062624	1030	0.00	1022	1020

European Union	371592	609	0.01	526	355
France	6670	81	0.02	74	66
Germany	9409	97	0.00	74	50
India	376162	613	0.02	578	595
Indonesia	7493	86	0.01	70	46
Italy	24717	157	0.05	153	136
Japan	114735	338	0.02	314	301
Mexico	22643	150	0.02	143	133
Russia	216277	465	0.02	457	443
SaudiArabia	24881	157	0.01	149	152
South Africa	19368	139	0.02	114	62
South Korea	19307	138	0.01	114	62
Turkey	2541	50	0.00	40	49
United Kingdom	30972	175	0.05	172	155
United States	2147402	1465	0.02	1382	1211

The G20 countries' CO_2 emissions are forecasted with varying degrees of accuracy by the linear regression model. Due to their complex and wide emissions patterns, China and the United States show the most significant errors. Much fewer errors are found in nations like Argentina and Turkey, indicating more consistent or predictable emissions trends. The error measures show higher emissions variability in nations like India, which may be the result of fast industrial and economic development. With its varied economies, the European Union displays a moderate inaccuracy, which is indicative of its combination of predictable and unpredictable emission patterns.

Random Forest:

Below plot shows the actual values vs predicted values in the dotted line for all countries:

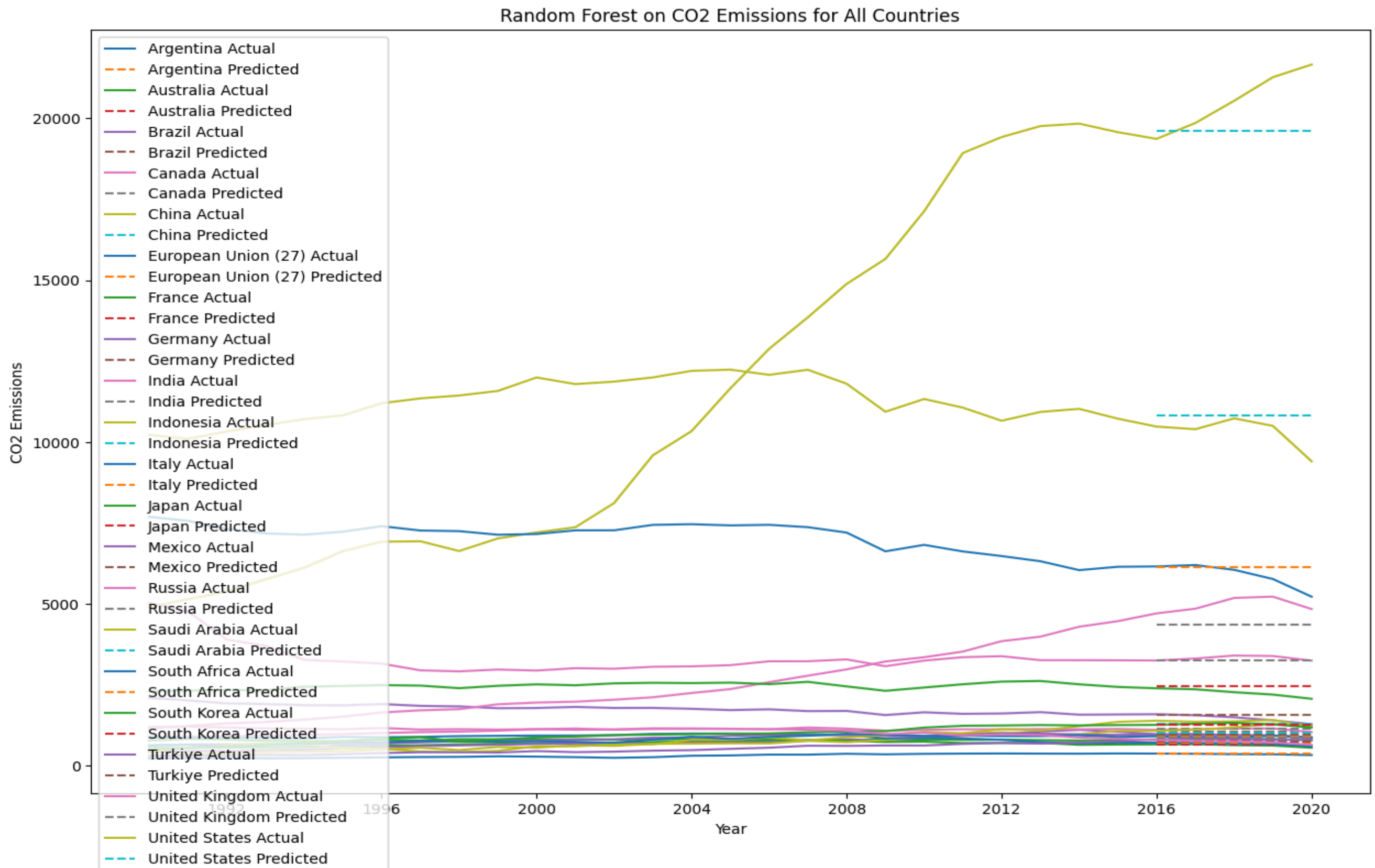


Table having different evaluation metrics for each country:

Country	MSE	RMSE	MSLE	MAE	MedAE
Argentina	647	25	0.00	19	19
Australia	733	27	0.00	24	29
Brazil	15414	124	0.02	118	116
Canada	1926	43	0.00	30	19
China	1565641	1251	0.00	1018	917

European Union	194555	441	0.01	289	81
France	2381	48	0.01	33	20
Germany	26848	163	0.01	121	88
India	394629	628	0.02	593	483
Indonesia	25919	161	0.02	140	149
Italy	2971	54	0.01	37	20
Japan	58447	241	0.01	211	198
Mexico	1723	41	0.00	33	27
Russia	8429	91	0.00	70	55
SaudiArabia	7975	89	0.00	87	78
South Africa	932	30	0.00	27	31
South Korea	2700	51	0.00	46	45
Turkey	6579	81	0.01	78	74
United Kingdom	17098	130	0.03	120	104
United States	488979	699	0.00	524	349

The predicting accuracy of CO2 emissions across the G20 countries by the random forest model demonstrates a wide range of performance. Given the complexity and possible variability of their emission patterns, the United States and China, two of the biggest polluters, have higher error metrics. Conversely, nations with lower errors include Canada, Argentina, and Mexico, suggesting more steady or predictable emission trends. The minor errors made by the European Union also reflect its diverse economic environment. Although the random forest model generally worked well, the accuracy levels varied due to the differing sizes and levels of complexity of the countries.

LSTM Model:

Below plot shows the actual values vs predicted values in the dotted line for all countries:

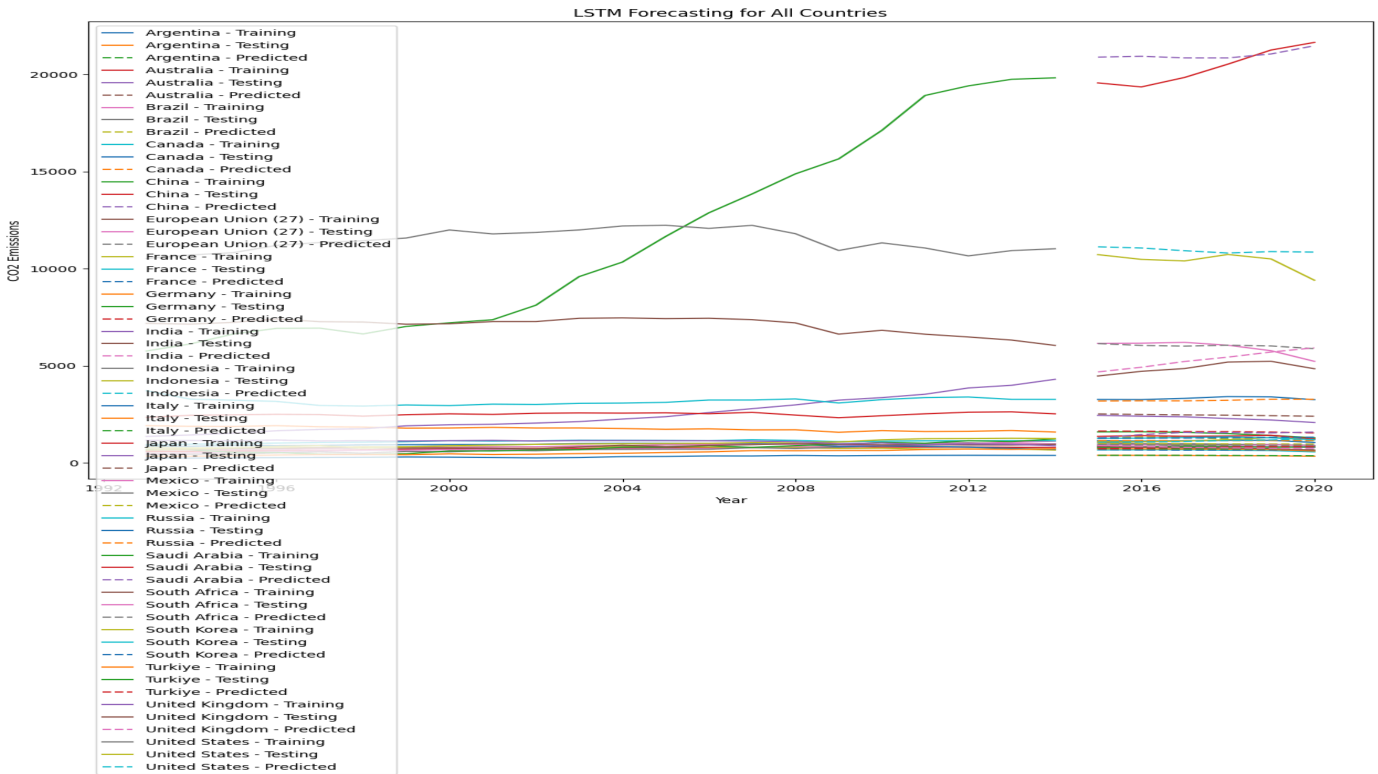


Table having different evaluation metrics for each country:

Country	MSE	RMSE	MSLE	MAE	MedAE
Argentina	190	13	0.00	11	8
Australia	335	18	0.00	16	16
Brazil	76920	227	0.06	272	265
Canada	2071	45	0.00	28	14
China	908850	953	0.00	768	663
European Union	89233	298	0.00	202	154

France	1022	31	0.00	23	14
Germany	17811	133	0.01	106	73
India	278846	528	0.01	432	308
Indonesia	3396	58	0.00	42	26
Italy	1430	37	0.00	23	11
Japan	34649	186	0.01	164	136
Mexico	2142	46	0.00	42	35
Russia	13070	114	0.00	100	101
SaudiArabia	37816	194	0.02	167	150
South Africa	1089	33	0.00	30	36
South Korea	3476	58	0.00	40	28
Turkey	2440	49	0.00	43	35
United Kingdom	1727	41	0.00	35	34
United States	505004	710	0.00	569	464

The G20 countries' CO_2 emissions forecasting outcomes from the LSTM model vary significantly. It is not unusual that China and the US have the most faults considering their huge economies and complex emission patterns. Brazil's high error metrics are indicative of possible emissions fluctuation. Conversely, countries such as France and Argentina have substantially smaller errors, suggesting more consistent and predictable trends in their emissions. These findings show the variable predicting accuracy levels for various nations, emphasizing the significance of adopting customized models for each nation's distinct features and emission patterns. The LSTM model did quite well overall, with most countries exhibiting minor errors.

6. Models Comparison and Analysis

In the process of forecasting CO2 emissions for the G20 countries, several models were employed, including ARIMA, linear regression, random forest, Holt-Winters, and LSTM. Each of these models provided insights into the emissions patterns, but their performance varied significantly across different countries.

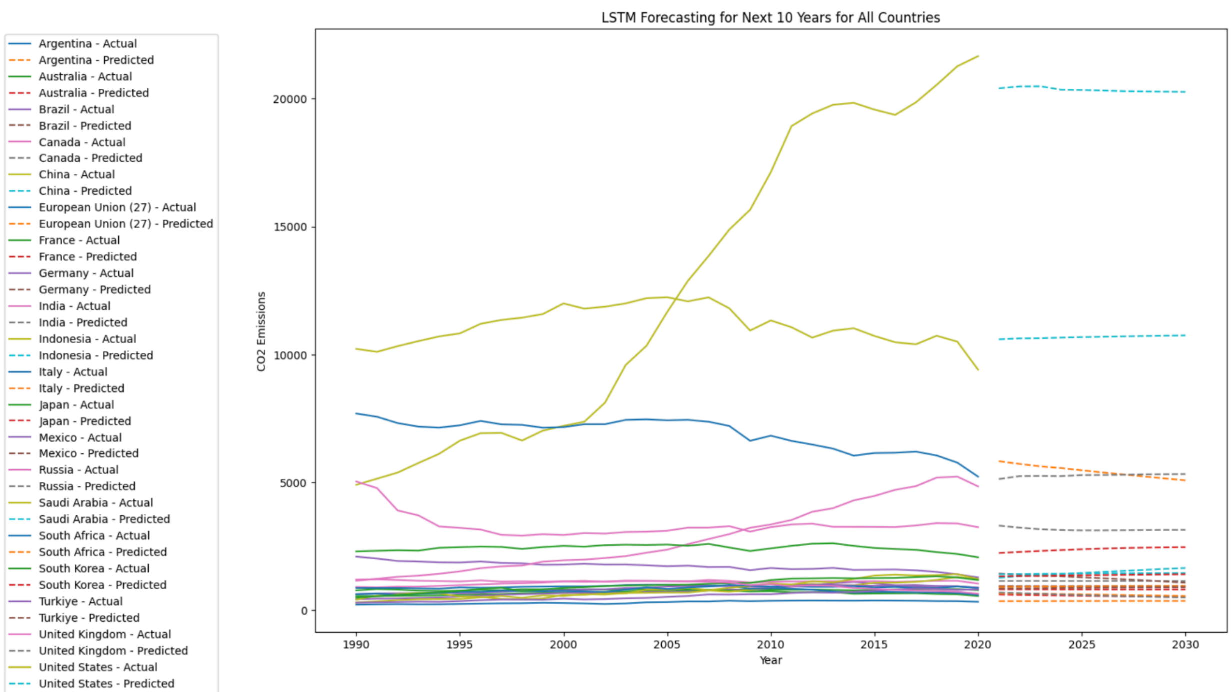
For nations with relatively low RMSE values and steady or predictable emission trends, such as Italy and the United Kingdom, the ARIMA model generally performed well. But in bigger, more complicated nations like China and the United States, where the RMSE values were higher than 500 and 1500, respectively, it had difficulty.

For a number of countries, the linear regression model outperformed ARIMA in terms of results. It still had issues with larger economies, where the emission patterns were less predictable, although it displayed lower RMSE values for nations like Germany and Canada. Despite its robustness, the random forest model performed inconsistently, with higher RMSE for nations like China and Brazil and smaller errors for nations like Argentina and Mexico.

The Holt-Winters model performed differently. It provided reasonably low RMSE values for certain nations, like France and Italy, but a much larger error for others, including China and Brazil, highlighting challenges in predicting their complex emission patterns.

On the other hand, the majority of countries consistently showed that the LSTM model performed better. Compared to the other models, the RMSE values for nations such as Brazil, China, and the US were lower. For the United States, for instance, the RMSE of the LSTM model was approximately 710, much less than the RMSE of the ARIMA model, which was approximately 540. The LSTM model demonstrated low RMSE values for countries such as Italy and Argentina, indicating its strong performance in smaller nations as well.

Predicting CO2 emissions of next 10 years till 2030 using the data till 2020 as training data using LSTM Model:



7. Conclusion

The results of this study shows that, when it comes to forecasting accuracy across different countries, the LSTM model performed better than the other models. All in all, the reduced RMSE values suggest that LSTM is a more appropriate model for capturing the intricate emission patterns of large and small economies. This implies that the LSTM model is a better option for predicting CO2 emissions in the G20 countries since it can better adjust to the distinct features of each nation. Because LSTM can identify long-term dependencies and capture nonlinear interactions, it has been shown to be useful in addressing the dynamic and diversified character of CO2 emissions data, producing forecasts that are more accurate and dependable.

References:

[1] Kumari, S., & Singh, S. K. (2023). Machine learning-based time series models for effective CO2 emission prediction in India. Environmental Science and Pollution Research, 30, 116601-116616. https://doi.org/10.1007/s11356-022-21723-8

[2] M. Malecki, "How to select a model for time series prediction," Neptune.ai, Mar. 20, 2023.[Online]. Available:https://neptune.ai/blog/select-model-for-time-series-prediction-task