I've used several different models for this dataset. Namely:

1. Using a Simple Moving Average
2. Linear Regression
3. Gradient Boosting Regression(with hyperparameter tuning)
4. Random Forest Regression(with hyperparameter tuning)
5. Voting Regressor

SIMPLE MOVING AVERAGE

Uses a moving average of CO2 emissions of all years of respective countries and calculates its mean. If there are any countries with less than 3 years of data it calculates the mean with available data.
The baseline prediction is this calculated value while the actual values are the values of the last 3 years, using which MAE(most absolute error) and

```
RMSE: 3.17
MAE: 2.41
```

LINEAR REGRESSION

Uses all the parameters(except the year and including the one-hot encoding of all the countries) and the CO2 emissions as the target value. The training set uses 80% of the actual values whereas testing 20%, after which it calculates the MAE and RMSE.

```
MAE: 4.90
RMSE: 5.68
```

GRADIENT BOOSTING REGRESSION

Uses a gradient boosting model which uses 200 decision trees and a learning rate of 0.01, to calculate the values of CO2 emissions. These are then compared with values and MAE and RMSE is calculated. This model is then optimized using hyperparameter tuning, with 5-fold cross validation, scored using negative mean squared error.

```
MAE: 4.97
RMSE: 5.76
```

	Feature	Importance
0	Avg Temperature (°C)	0.141665
6	Population	0.141468
3	Renewable Energy (%)	0.136188
2	Rainfall (mm)	0.132682
5	Forest Area (%)	0.127663
1	Sea Level Rise (mm)	0.112285
4	Extreme Weather Events	0.083166

```
17          Country_Russia    0.012530
19              Country_UK    0.010972
20             Country_USA    0.010829
7        Country_Australia    0.010495
18    Country_South Africa    0.010263
12         Country_Germany    0.009462
13           Country_India    0.008576
10           Country_China    0.008235
14       Country_Indonesia    0.008090
11          Country_France    0.007670
16          Country_Mexico    0.007589
15           Country_Japan    0.007139
8           Country_Brazil    0.006902
9           Country_Canada    0.006130
```

RANDOM FOREST REGRESSION

Uses a Random Forest Regressor which uses 200 decision trees, to calculate the values of CO2 which are then compared with actual CO2 values. It then calculates the MAE and RMSE. This model is then optimized using hyperparameter tuning, with 3-fold cross validation and scored using negative mean squared error.

```
MAE: 5.03
RMSE: 5.89
```

VOTING REGRESSOR

Uses a linear regression, gradient boosting and random forest model to assign weight to values based on performance and then averaged out to approximate the influence of each of the values. It then compares this data to the actual values and calculates the RMSE and MAE.

An important thing to note is that I've also used the feature_importance_ library which calculates the influence of each input variable to the output variable, using this Gini Index, which hence allocates weight to the input variable.

	Feature	Importance
0	Avg Temperature (°C)	0.141665
6	Population	0.141468
3	Renewable Energy (%)	0.136188
2	Rainfall (mm)	0.132682
5	Forest Area (%)	0.127663
1	Sea Level Rise (mm)	0.112285
4	Extreme Weather Events	0.083166
17	Country_Russia	0.012530
19	Country_UK	0.010972
20	Country_USA	0.010829
7	Country_Australia	0.010495
18	Country_South Africa	0.010263
12	Country_Germany	0.009462
13	Country_India	0.008576
10	Country_China	0.008235
14	Country_Indonesia	0.008090
11	Country_France	0.007670
16	Country_Mexico	0.007589
15	Country_Japan	0.007139
8	Country_Brazil	0.006902
9	Country_Canada	0.006130

```
Tuned MAE: 4.93
Tuned RMSE: 5.70
```