Assignment

The World Bank hosts a wide range of data sets and variables. Go to the website https://databank.worldbank.org/source/world-development-indicators .

From the World Development Indicators download the following variables and store in csv format in a file on your computer:

CO2 emissions (metric tons per capita)
GDP per capita growth (annual)
GDP per capita (constant 2015 US dollar)
Forest area (sq. km)
Livestock production index
Electricity production from renewable sources excluding hydroelectric (kWh)
Energy use (kg of oil equivalent per capita)
Renewable energy consumption (of total final energy consumption)
Electric power consumption
Population growth (annual)
Urban population growth (annual)
Population total

Your task is to fit a machine learning model for the relationship between CO2 emissions and further explanatory factors.
Use numpy and pandas, as well as scikit-learn. Load the data into a Jupyter Notebook environment. Split your data set into training and test data samples. Fit a linear regression, a decision tree and random forests model. Display the regression coefficients from the linear regression. Display performance metrics for your models. Which model would you preferably choose?