SAMPLE EXAMPLE OF RANDOM FOREST ALGORITHM:

CODE:

```
#The problem here is to predict the gas consumption (in millions of gallons) in 48 of the US s
tates based on petrol tax (in cents),
#per capita income (dollars), paved highways (in miles) and the proportion of population with
 the driving license.
import pandas as pd
import numpy as np
dataset = pd.read_csv('petrol_consumption.csv')
dataset.head()
dataset.tail()
#Preparing Data for training
X = dataset.iloc[:, 0:4].values
y = dataset.iloc[:, 4].values
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)


#Feature Scaling
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)


#Training
from sklearn.ensemble import RandomForestRegressor

regressor = RandomForestRegressor(n_estimators=20, random_state=0)
regressor.fit(X_train, y_train)
y_pred = regressor.predict(X_test)
```

OUTPUT:

	Petrol_tax	Average_income	Paved_Highways	Population_Driver_licence(%)	Petrol_Consumption
0	9.0	3571	1976	0.525	541
1	9.0	4092	1250	0.572	524
2	9.0	3865	1586	0.580	561
3	7.5	4870	2351	0.529	414
4	8.0	4399	431	0.544	410

Data set glimpses

	Petrol_tax	Average_income	Paved_Highways	Population_Driver_licence(%)	Petrol_Consumption
43	7.0	3745	2611	0.508	591
44	6.0	5215	2302	0.672	782
45	9.0	4476	3942	0.571	510
46	7.0	4296	4083	0.623	610
47	7.0	5002	9794	0.593	524

Data set glimpses

```
Mean Absolute Error: 51.76500000000001
Mean Squared Error: 4216.166749999999
Root Mean Squared Error: 64.93201637097064
```

RMSE with 20 Trees

```
Mean Absolute Error: 49.304000000000016
Mean Squared Error: 3758.844960000001
Root Mean Squared Error: 61.30941983088733
```

RMSE with 50 Trees

```
Mean Absolute Error: 48.33899999999999
Mean Squared Error: 3494.2330150000003
Root Mean Squared Error: 59.112037818028234
```

RMSE with 200 Trees

```
Mean Absolute Error: 48.17666666666667
Mean Squared Error: 3472.820688888889
Root Mean Squared Error: 58.93064303814179
```

RMSE with 300 Trees

You can see that the error values decreases with the increase in number of estimator. After 200 the rate of decrease in error diminishes, so therefore 200 is a good number for n_estimators