In [1]:
```
import pandas as pd
import numpy as np

from sklearn.model_selection import train_test_split

from sklearn.ensemble import RandomForestRegressor
```

In [ ]:
```
# read the empty column

fill = pd.read_csv("datasets/training_data.csv", skiprows = 4933, nrows = 38,
```

In [34]:
```
# define the input variable
train_input = df[['Average_OAT', 'Humidity', 'UV_Index', 'Average_Rainfall']]
train_output = df.CoolingLoad

fill_input = fill[['Average_OAT', 'Humidity', 'UV_Index', 'Average_Rainfall']]
```

In [35]:
```
# model training
model = RandomForestRegressor(n_estimators = 10000, max_features = 4)
model.fit(train_input, train_output)
```

Out[35]:
```
RandomForestRegressor(max_features=4, n_estimators=10000)
```

In [36]:
```
# Using the model to fill data
predictions = model.predict(fill_input)
prediction = pd.DataFrame(predictions, columns=['predictions']).to_csv('predi
pd.DataFrame(predictions)
```

Out[36]:

	0
0	3134.212634
1	3075.216591
2	3164.287387
3	3227.526100
4	3169.132431
5	3083.266315
6	3193.232186
7	3210.443660
8	3135.811745
9	3152.727874
10	3142.914584

In [ ]: