Splitting the variables

In [ ]:
```
x = data.drop('mpg',axis=1)
y = data.mpg
```

Splitting the data

In [29]:
```
from sklearn.model_selection import train_test_split
```

In [30]:
```
trainx,testx,trainy,testy = train_test_split(x,y,test_size=0.2)
```

In [31]:
```
testx
```

Out[31]:

	cylinders	displacement	horsepower	weight	acceleration	model year	origin	Brand
25	8	360.0	215	4615	14.0	70	1	11
236	4	140.0	89	2755	15.8	77	1	11
55	4	97.0	60	1834	19.0	71	2	29
337	4	107.0	72	2290	17.0	80	3	13
173	4	119.0	97	2545	17.0	75	3	8
...	...	...	...	...	...	...	...	...
227	6	225.0	100	3630	17.7	77	1	22
357	4	119.0	100	2615	14.8	81	3	8
152	6	225.0	95	3264	16.0	75	1	22
170	4	140.0	78	2592	18.5	75	1	23
5	8	429.0	198	4341	10.0	70	1	11

80 rows × 8 columns

Model Training

In [4]:
```
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import f1_score
```

In [40]:
```
model = RandomForestRegressor().fit(trainx,trainy)
```

Model Evaluation Metrics

In [41]:
```
model.score(testx,testy)
```

Out[41]:
```
0.8717371401824269
```

In [42]:
```
predy=model.predict(testx)
predy
```

Out[42]:
```
array([11.74 , 24.331, 27.82 , 33.531, 23.597, 37.68 , 23.153, 19.965,
       33.104, 20.477, 33.793, 23.056, 16.25 , 19.805, 17.409, 14.43 ,
       13.24 , 28.784, 17.889, 23.96 , 15.577, 18.638, 24.708, 28.958,
       14.105, 22.984, 24.663, 19.249, 24.794, 25.22 , 23.077, 36.332,
       23.281, 14.82 , 37.313, 31.87 , 35.178, 14.415, 26.64 , 19.662,
       15.175, 19.298, 25.534, 25.915, 18.76 , 14.515, 20.42 , 29.461,
       19.751, 14.365, 31.583, 14.285, 22.496, 16.079, 12.02 , 22.627,
       31.831, 27.403, 14.56 , 21.015, 21.343, 16.282, 31.895, 24.403,
       30.58 , 19.205, 24.155, 24.01 , 27.075, 31.952, 24.98 , 19.633,
       38.218, 28.839, 24.479, 19.083, 27.737, 19.022, 23.472, 13.59 ])
```

In [6]:
```
f1_score(testy,predy)
```

```
0.87239871267
```

In [84]:
```
file = open('binary.pkl','wb')
pickle.dump(model,file)
```

In [85]:
```
file.close()
```