In [27]:
```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier

data = pd.read_csv('C:/Users/USER/Desktop/MLENSEMBLESDOCS-02NOV2021/mnist/mnist_train.csv')
```

In [28]:
```
df_X = data.iloc[:, 1:]

df_Y = data.iloc[:, 0]
```

In [29]:
```
X_train, X_test, Y_train, Y_test = train_test_split(df_X, df_Y, test_size = 0.2, random_state = 4)
```

In [30]:
```
data.shape
```

Out[30]:
```
(60000, 785)
```

In [31]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60000 entries, 0 to 59999
Columns: 785 entries, label to 28x28
dtypes: int64(785)
memory usage: 359.3 MB
```

In [32]:
```
#Decision Tree
dt = DecisionTreeClassifier()
dt.fit(X_train, Y_train)
```

Out[32]:
```
DecisionTreeClassifier()
```

In [34]:
```
dt.score(X_test, Y_test)
```

Out[34]:
```
0.8655833333333334
```

In [35]:
```
#Testing for Model overfit on train data
dt.score(X_train, Y_train)

#A score of 1.0 implies that the model is 100% overfit to the training data.
```

```
1.0
```

Out[35]:

In [33]:
```
from sklearn.ensemble import AdaBoostClassifier
from sklearn.preprocessing import LabelEncoder
from sklearn.tree import DecisionTreeClassifier
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import metrics
```