

SPLITTING DATA

Splitting data into independent and dependent variables

Identifying Independent & dependent variables:

In this activity, the dependent and independent variables are to be identified. The last column (Result) in the dataset is the dependent variable which is dependent on the 30 different factors. The independent columns are considered as x and the dependent column as y

INPUT THE COMMAND AS SHOWN

```
#Splitting data into train and test
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)
```

Splitting the data:

After identifying the dependent and independent variables, the dataset now has to be split into two sets, one set is used for training the model and the second set is used for testing how good the model is built. The split ratio we consider is 80% for training and 20% for testing.

INPUT THE COMMAND AS SHOWN

```
#Splitting data into train and test
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)
```