

```
import pandas as pd
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier

data=pd.read_csv('/content/drive/MyDrive/ml lab/daily_weather.csv')

data.head()
```

	number	air_pressure_9am	air_temp_9am	avg_wind_direction_9am	avg_wind
0	0	918.060000	74.822000	271.100000	
1	1	917.347688	71.403843	101.935179	
2	2	923.040000	60.638000	51.000000	
3	3	920.502751	70.138895	198.832133	
4	4	921.160000	44.294000	277.800000	

```
data.shape

(1095, 11)
```

```
data.info
```

```
<bound method DataFrame.info of
avg_wind_direction_9am \
0      0      918.060000      74.822000      271.100000
1      1      917.347688      71.403843      101.935179
2      2      923.040000      60.638000      51.000000
3      3      920.502751      70.138895      198.832133
4      4      921.160000      44.294000      277.800000
...
1090   1090      918.900000      63.104000      192.900000
1091   1091      918.710000      49.568000      241.600000
1092   1092      916.600000      71.096000      189.300000
1093   1093      912.600000      58.406000      172.700000
1094   1094      921.530000      77.702000      97.100000
...
      avg_wind_speed_9am  max_wind_direction_9am  max_wind_speed_9am \
0      2.080354      295.400000      2.863283
1      2.443009      140.471548      3.533324
2      17.067852      63.700000      22.100967
3      4.337363      211.203341      5.190045
4      1.856660      136.500000      2.863283
...
1090      3.869906      207.300000      5.212070
1091      1.811921      227.400000      2.371156
1092      3.064608      200.800000      3.892276
1093      3.825167      189.100000      4.764682
```

```
1094          3.265932          125.900000          4.451511
```

```

rain_accumulation_9am  rain_duration_9am  relative_humidity_9am  \
0          0.0          0.0          42.420000
1          0.0          0.0          24.328697
2          0.0          20.0          8.900000
3          0.0          0.0          12.189102
4          8.9          14730.0          92.410000
...          ...          ...          ...
1090         0.0          0.0          26.020000
1091         0.0          0.0          90.350000
1092         0.0          0.0          45.590000
1093         0.0          0.0          64.840000
1094         0.0          0.0          14.560000

```

```

relative_humidity_3pm
0          36.160000
1          19.426597
2          14.460000
3          12.742547
4          76.740000
...          ...
1090         38.180000
1091         73.340000
1092         52.310000
1093         58.280000
1094         15.100000

```

```
[1095 rows x 11 columns]>
```

```
data.columns
```

```

Index(['number', 'air_pressure_9am', 'air_temp_9am',
      'avg_wind_direction_9am',
      'avg_wind_speed_9am', 'max_wind_direction_9am', 'max_wind_speed_9am',
      'rain_accumulation_9am', 'rain_duration_9am',
      'relative_humidity_9am',
      'relative_humidity_3pm'],
      dtype='object')

```

```
data[data.isnull().any(axis=1)].head(500)
```

	number	air_pressure_9am	air_temp_9am	avg_wind_direction_9am	avg_w:
	16	16	917.890000	NaN	169.200000
	111	111	915.290000	58.820000	182.600000
	177	177	915.900000	NaN	183.300000
	262	262	923.596607	58.380598	47.737753
	277	277	920.480000	62.600000	194.400000
	334	334	916.230000	75.740000	149.100000
	358	358	917.440000	58.514000	55.100000
	361	361	920.444946	65.801845	49.823346
	381	381	918.480000	66.542000	90.900000
	409	409	NaN	67.853833	65.880616
	517	517	920.570000	53.600000	100.100000
	519	519	916.250000	55.670000	176.400000
	546	546	NaN	42.746000	251.100000
	620	620	921.200000	56.786000	192.300000
	625	625	912.400000	50.774000	171.600000
	656	656	920.830000	66.344000	NaN
	670	670	910.920000	48.362000	156.500000
	672	672	922.448945	72.863773	NaN
	705	705	911.900000	59.072000	199.800000
	731	731	922.970166	51.391847	33.810942
	737	737	917.895130	76.804690	104.771020
	788	788	917.923442	73.249717	42.101739
	840	840	918.043767	NaN	181.774042
	848	848	915.850000	67.560000	246.500000

```
del data['number']
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1095 entries, 0 to 1094
Data columns (total 10 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   air_pressure_9am                      1092 non-null   float64
1   air_temp_9am                          1090 non-null   float64
2   avg_wind_direction_9am                1091 non-null   float64
3   avg_wind_speed_9am                    1092 non-null   float64
4   max_wind_direction_9am                1092 non-null   float64
```

```

5    max_wind_speed_9am    1091 non-null    float64
6    rain_accumulation_9am 1089 non-null    float64
7    rain_duration_9am    1092 non-null    float64
8    relative_humidity_9am 1095 non-null    float64
9    relative_humidity_3pm 1095 non-null    float64
dtypes: float64(10)
memory usage: 85.7 KB

```

```
data=data.dropna()
```

```
data.shape
```

```
(1064, 10)
```

```
data.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1064 entries, 0 to 1094
Data columns (total 10 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   air_pressure_9am                     1064 non-null   float64
1   air_temp_9am                         1064 non-null   float64
2   avg_wind_direction_9am               1064 non-null   float64
3   avg_wind_speed_9am                   1064 non-null   float64
4   max_wind_direction_9am               1064 non-null   float64
5   max_wind_speed_9am                   1064 non-null   float64
6   rain_accumulation_9am                 1064 non-null   float64
7   rain_duration_9am                    1064 non-null   float64
8   relative_humidity_9am                 1064 non-null   float64
9   relative_humidity_3pm                 1064 non-null   float64
dtypes: float64(10)
memory usage: 91.4 KB

```

```
clean_data=data.copy()
```

```
clean_data['relative_humidity_3pm'].head()
```

```

0    36.160000
1    19.426597
2    14.460000
3    12.742547
4    76.740000
Name: relative_humidity_3pm, dtype: float64

```

```

clean_data['high_humidity_label'] = (clean_data['relative_humidity_3pm']>28)
clean_data['high_humidity_label'].head()



```

```

0    1
1    0
2    0
3    0
4    1
Name: high_humidity_label, dtype: int64

```

```
y=clean_data[['high_humidity_label']].copy()
y.head()
```

high_humidity_label		
0	1	
1	0	
2	0	
3	0	
4	1	

```
clean_data.columns
```

```
Index(['air_pressure_9am', 'air_temp_9am', 'avg_wind_direction_9am',
      'avg_wind_speed_9am', 'max_wind_direction_9am', 'max_wind_speed_9am',
      'rain_accumulation_9am', 'rain_duration_9am',
      'relative_humidity_9am',
      'relative_humidity_3pm', 'high_humidity_label'],
      dtype='object')
```

```
morning_features=['air_pressure_9am', 'air_temp_9am', 'avg_wind_direction_9am',
                  'avg_wind_speed_9am', 'max_wind_direction_9am', 'max_wind_speed_9am',
                  'rain_accumulation_9am', 'rain_duration_9am', 'relative_humidity_9am']
```

```
x=clean_data[morning_features].copy()
x.columns
```


```
Index(['air_pressure_9am', 'air_temp_9am', 'avg_wind_direction_9am',
      'avg_wind_speed_9am', 'max_wind_direction_9am', 'max_wind_speed_9am',
      'rain_accumulation_9am', 'rain_duration_9am',
      'relative_humidity_9am'],
      dtype='object')
```

```
y.columns
```

```
Index(['high_humidity_label'], dtype='object')
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.33,random_state=42)
```

```
humidity_classifier=DecisionTreeClassifier(max_leaf_nodes=10, random_state=0)
humidity_classifier.fit(x_train,y_train)
```


DecisionTreeClassifier
 DecisionTreeClassifier(max_leaf_nodes=10, random_state=0)

```
y_predicted=humidity_classifier.predict(x_test)
```

```
accuracy_score(y_test,y_predicted)
```

```
0.8892045454545454
```

```
confusion_matrix(y_test,y_predicted)
```

```
array([[166, 14],  
       [ 25, 147]])
```

```
from sklearn.tree import plot_tree
```

```
from matplotlib.pyplot import rcParams  
rcParams['figure.figsize'] = 80,50  
plot_tree(humidity_classifier)
```

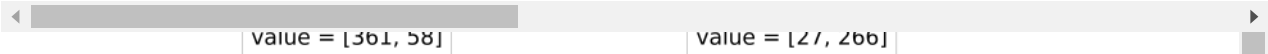
```
[Text(0.45454545454545453, 0.9166666666666666, 'x[8] <= 28.364\ngini = 0.496\nsamples = 712\nvalue = [388, 324]'),
Text(0.2727272727272727, 0.75, 'x[8] <= 19.955\ngini = 0.239\nsamples = 419\nvalue = [361, 58]'),
Text(0.18181818181818182, 0.5833333333333334, 'gini = 0.103\nsamples = 311\nvalue = [294, 17]'),
Text(0.36363636363636365, 0.5833333333333334, 'x[0] <= 917.284\ngini = 0.471\nsamples = 108\nvalue = [67, 41]'),
Text(0.2727272727272727, 0.4166666666666667, 'x[4] <= 130.6\ngini = 0.432\nsamples = 38\nvalue = [12, 26]'),
Text(0.18181818181818182, 0.25, 'x[4] <= 72.8\ngini = 0.48\nsamples = 15\nvalue = [9, 6]'),
Text(0.09090909090909091, 0.08333333333333333, 'gini = 0.444\nsamples = 9\nvalue = [3, 6]'),
Text(0.2727272727272727, 0.08333333333333333, 'gini = 0.0\nsamples = 6\nvalue = [6, 0]'),
Text(0.36363636363636365, 0.25, 'gini = 0.227\nsamples = 23\nvalue = [3, 20]'),
Text(0.45454545454545453, 0.4166666666666667, 'gini = 0.337\nsamples = 70\nvalue = [55, 15]'),
```

```
import matplotlib.pyplot as plt
%matplotlib inline
```

```
df = pd.read_csv('/content/drive/MyDrive/ml_lab/MNIST_data.csv')
df.head()
```

	label	pixel0	pixel1	pixel2	pixel3	pixel4	pixel5	pixel6	pixel7	pixel8
0	1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	4	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0

5 rows × 785 columns



```
df.shape
```

(42000, 785)



```
df.info
```

<bound method DataFrame.info of										
pixel4	pixel5	pixel6	pixel7	pixel8	label	pixel0	pixel1	pixel2	pixel3	pixel4
0	1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	4	0	0	0	0	0	0	0	0	0

26/09/2023, 16:36

Untitled3.ipynb - Colaboratory

4	0	0	0	0	0	0	0	0	0
...
41995	0	0	0	0	0	0	0	0	0
41996	1	0	0	0	0	0	0	0	0
41997	7	0	0	0	0	0	0	0	0
41998	6	0	0	0	0	0	0	0	0
41999	9	0	0	0	0	0	0	0	0

	pixel8	...	pixel774	pixel775	pixel776	pixel777	pixel778	\
0	0	...	0	0	0	0	0	
1	0	...	0	0	0	0	0	
2	0	...	0	0	0	0	0	
3	0	...	0	0	0	0	0	
4	0	...	0	0	0	0	0	
...	
41995	0	...	0	0	0	0	0	
41996	0	...	0	0	0	0	0	
41997	0	...	0	0	0	0	0	
41998	0	...	0	0	0	0	0	
41999	0	...	0	0	0	0	0	

	pixel779	pixel780	pixel781	pixel782	pixel783
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
...
41995	0	0	0	0	0
41996	0	0	0	0	0
41997	0	0	0	0	0
41998	0	0	0	0	0
41999	0	0	0	0	0

[42000 rows x 785 columns]>

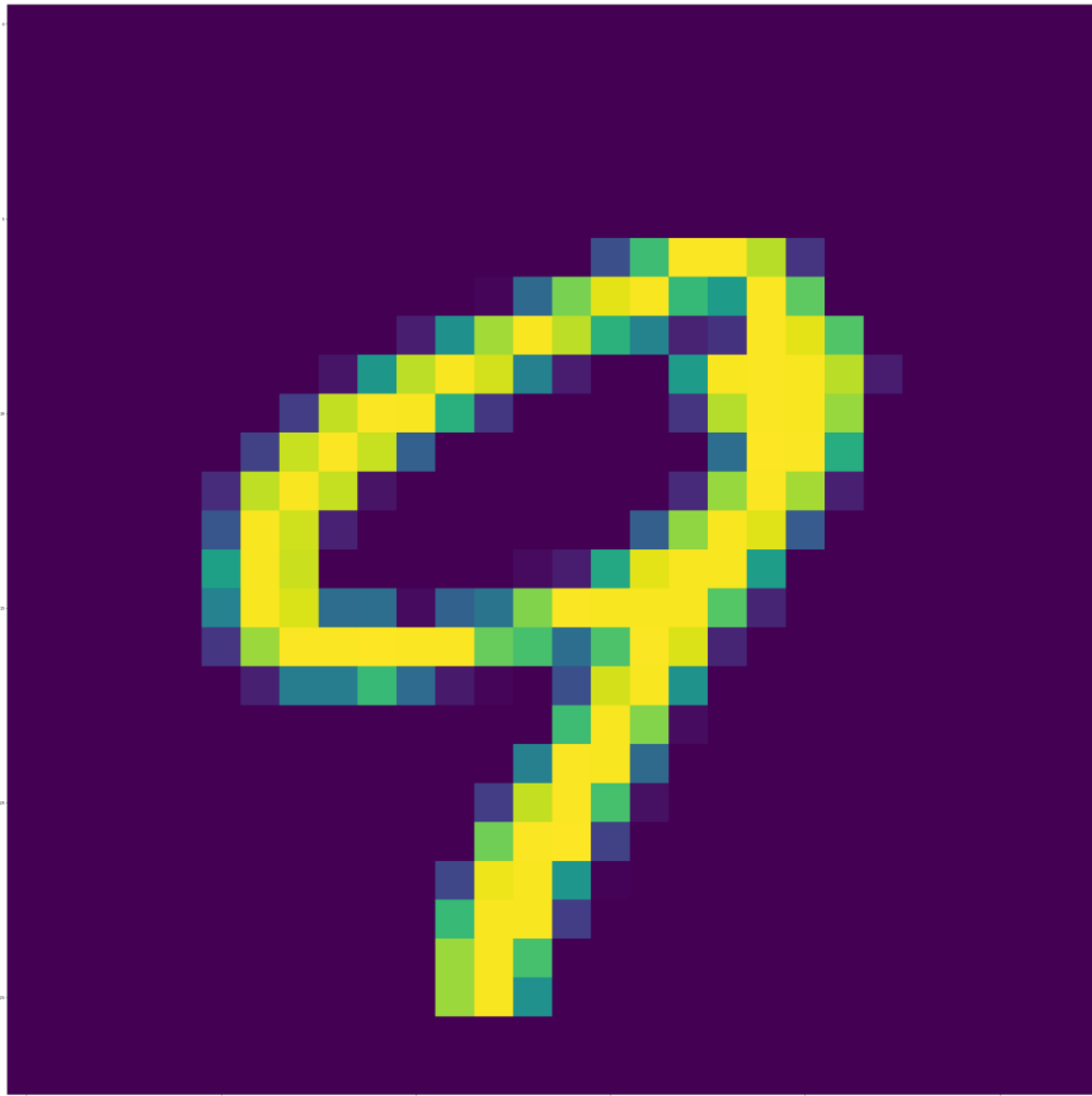


```
x=df.iloc[:,1:].values
x.shape

(42000, 784)

plt.imshow(x[100].reshape(28,28))
```


<matplotlib.image.AxesImage at 0x7e56c1c3d30>



```
y=df.iloc[:,0].values  
y.shape
```

```
(42000,)
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=42)
```

```
x_train.shape
```

```
(33600, 784)
```

```
classifier = DecisionTreeClassifier()
```

```
classifier.fit(x_train,y_train)
```

```
▼ DecisionTreeClassifier
```

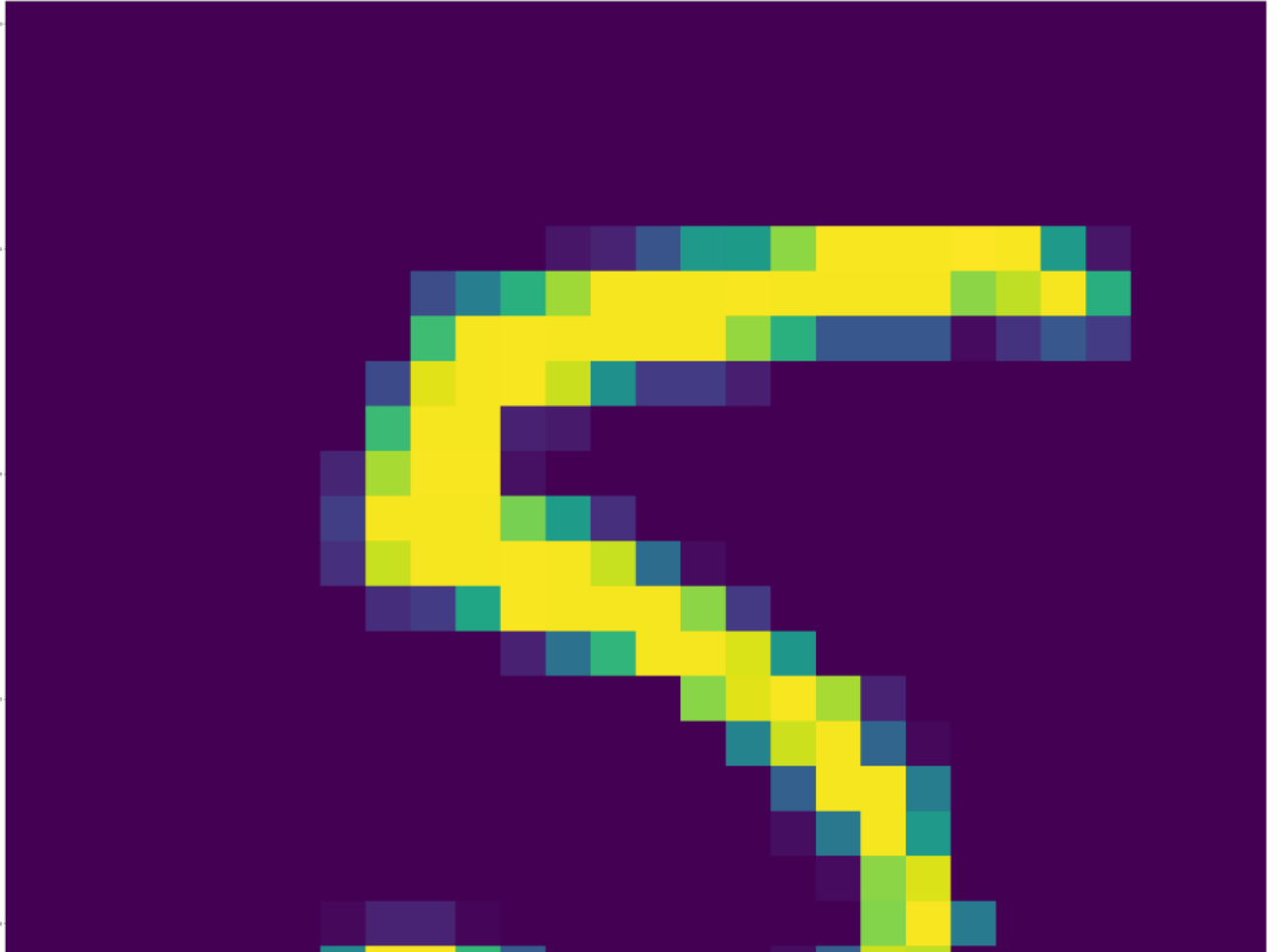
```
DecisionTreeClassifier()
```

```
y_test[100]
```

```
5
```

```
plt.imshow(x_test[100].reshape(28,28))
```

```
<matplotlib.image.AxesImage at 0x7e56ca177880>
```



```
classifier.predict(x_test[100].reshape(1,784))
```

```
array([5])
```



```
y_test[812]
```

```
1
```

```
plt.imshow(x_test[99].reshape(28,28))
```

```
classifier.predict(x_test[812].reshape(1,784))
```

```
array([1])
```

```
y_predicted = classifier.predict(x_test)
```

```
accuracy_score(y_test,y_predicted)
```

```
0.851547619047619
```

```
confusion_matrix(y_test,y_predicted)
```

```
array([[745,  0,  8, 10,  3, 10,  9,  6,  7,  3],
       [ 0, 892, 19,  9,  9, 10,  1,  9, 17,  2],
```

```
[ 10, 16, 683, 30, 8, 2, 13, 31, 26, 10],  
[ 6, 10, 22, 722, 10, 41, 4, 18, 31, 21],  
[ 6, 9, 16, 8, 684, 11, 17, 6, 19, 32],  
[ 17, 9, 7, 39, 12, 611, 23, 12, 19, 18],  
[ 11, 0, 14, 9, 8, 25, 706, 7, 17, 6],  
[ 1, 3, 22, 9, 11, 5, 2, 768, 14, 20],  
[ 16, 18, 24, 30, 15, 36, 18, 11, 655, 12],  
[ 5, 9, 9, 16, 48, 22, 3, 32, 18, 687]])
```