

```
In [3]: #importing dependencies
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
%matplotlib inline
```

```
In [4]: #read
data = pd.read_csv("mnist_test.csv")
data = pd.read_csv("mnist_train.csv")
```

```
In [5]: data.head()
```

```
Out[5]:
```

	label	1x1	1x2	1x3	1x4	1x5	1x6	1x7	1x8	1x9	...	28x19	28x20	28x21	28x22	28x23
0	5	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	4	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4	9	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

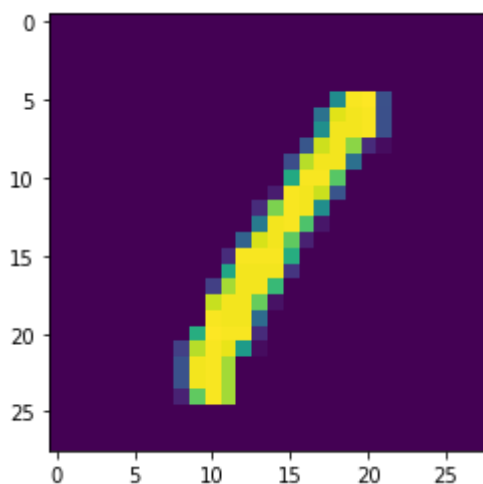
5 rows × 785 columns



```
In [6]: #extracting
a = data.iloc[3,1:].values
```

```
In [7]: #reshaping
a = a.reshape(28,28).astype('uint8')
plt.imshow(a)
```

```
Out[7]: <matplotlib.image.AxesImage at 0x295f768e0d0>
```



```
In [8]: df_x = data.iloc[:,1:]
df_y = data.iloc[:,0]
```

```
In [9]: x_train, x_test, y_train, y_test = train_test_split(df_x , df_y ,test_size=0.2 , ran
```

```
In [10]: y_train.head()
```

```
Out[10]: 20379    4
53032    4
27005    7
30510    8
508      1
Name: label, dtype: int64
```

```
In [16]: rf = RandomForestClassifier(n_estimators=100)
```

```
In [12]: rf.fit(x_train,y_train)
```

```
Out[12]: RandomForestClassifier()
```

```
In [13]: #prediction on test data
pred = rf.predict(x_test)
```

```
In [17]: pred
```

```
Out[17]: array([2, 7, 6, ..., 6, 4, 2], dtype=int64)
```

```
In [18]: #check prediction accuracy
s = y_test.values

count=0
for i in range (len(pred)):
    if pred [i] == s[i]:
        count = count+1
```

```
In [19]: count
```

```
Out[19]: 11594
```

```
In [20]: len(pred)
```

```
Out[20]: 12000
```

```
In [21]: #accuracy value
19594/12000
```

```
Out[21]: 1.6328333333333334
```

```
In [ ]:
```