Name : Syed Riaz Ali

Email : syedriazali1997@gmail.com

Whatsapp : 923002502513

Day_26 : Random Forest Classification (For Large dataset)

I have not use the 1.2 GB data because I don't have a powerful computing system. Instead of using 1.2 GB extra_32x32.mat data I used 184 MB train_32x32.mat data which is also a large data and it took about 15 mins to train this dataset.

In [1]:
```
# Importing Libraries

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

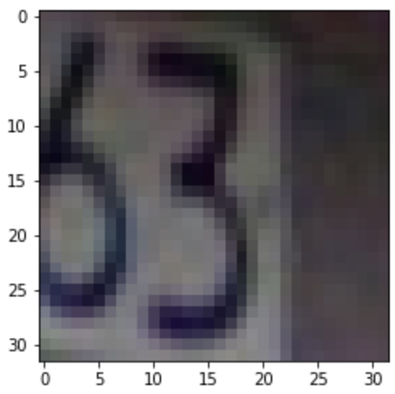
In [5]:
```
# Loading the data set

import scipy.io
df = scipy.io.loadmat('train_32x32.mat')
```

In [7]:
```
# Separating the input and output variables
X = df['X']
y = df['y']
```

In [10]:
```
# Viewing the elements in the dataset

img_index = 25
plt.imshow(X[:,:,:,img_index])
plt.show()
print(y[img_index])
```




```
[3]
```

In [11]:
```
# Reshaping the data set

from sklearn.utils import shuffle
X = X.reshape(X.shape[0]*X.shape[1]*X.shape[2],X.shape[3]).T
y = y.reshape(y.shape[0],)
X, y = shuffle(X, y, random_state=42)
```

In [12]:
```
# Applying Machine learning Algorithm
from sklearn.ensemble import RandomForestClassifier
clf = RandomForestClassifier()
print(clf)
```

```
RandomForestClassifier()
```

In [14]:
```
# Testing the model using train_test_split technique
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=42)
clf.fit(X_train, y_train)
```

Out[14]:
```
RandomForestClassifier()
```

In [15]:
```
# Accuracy Test
from sklearn.metrics import accuracy_score
preds = clf.predict(X_test)
print("Accuracy:", accuracy_score(y_test,preds))
```

```
Accuracy: 0.7055692055692055
```

In [ ]:
```

```