In [26]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

importing the dataset into workspace

In [27]:
```
data=pd.read_csv("ibm1.csv")
```

In [28]:
```
data
```

Out[28]:

	Date	Closing Value
0	02-01-1986	25.56
1	03-01-1986	26.00
2	06-01-1986	26.53
3	07-01-1986	25.85
4	08-01-1986	25.87
...	...	...
8218	03-07-2018	74.19
8219	04-07-2018	NaN
8220	05-07-2018	73.05
8221	06-07-2018	73.78
8222	09-07-2018	73.93

8223 rows × 2 columns

handling missing data

In [29]:
```
data.isnull().any()
```

Out[29]:
```
Date             False
Closing Value     True
dtype: bool
```

In [30]:
```
data.isnull().sum()
```

Out[30]:
```
Date             0
Closing Value    7
dtype: int64
```

In [31]:
```
data.dropna(axis=0,inplace=True)
```

In [32]:
```
data.isnull().sum()
```

Out[32]:
```
Date             0
Closing Value    0
dtype: int64
```

Feature scaling

In [33]:
```
data_oil=data.reset_index()['Closing Value']
data_oil
```

Out[33]:
```
0       25.56
1       26.00
2       26.53
3       25.85
4       25.87
        ...
8211    73.89
8212    74.19
8213    73.05
8214    73.78
8215    73.93
Name: Closing Value, Length: 8216, dtype: float64
```

In [34]:
```
from sklearn.preprocessing import MinMaxScaler
Scaler=MinMaxScaler(feature_range=(0,1))
data_oil=Scaler.fit_transform(np.array(data_oil).reshape(-1,1))
```

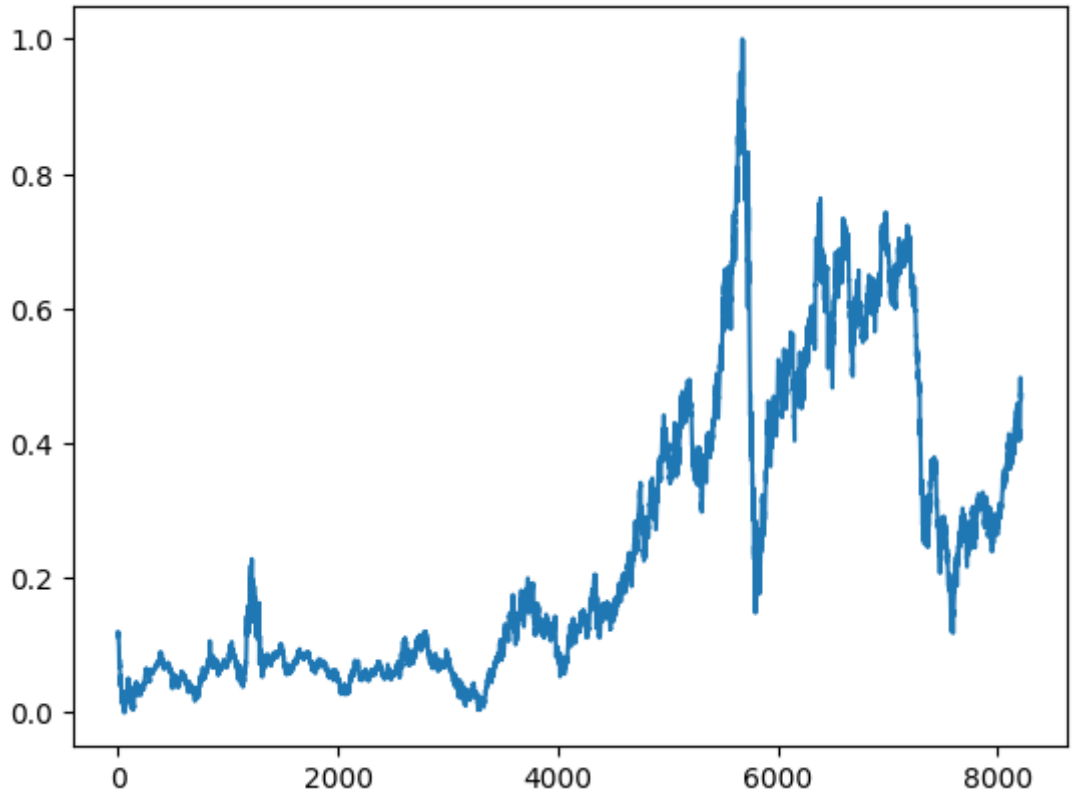
In [35]:
```
data_oil
```

Out[35]:
```
array([[0.11335703],
       [0.11661484],
       [0.12053902],
       ...,
       [0.46497853],
       [0.47038353],
       [0.47149415]])
```

Data visualization

In [36]:
```
plt.plot(data_oil)
```

Out[36]:
```
[<matplotlib.lines.Line2D at 0x1f9741ee290>]
```



spliting the data into train and test

In [37]:
```
training_size=int(len(data_oil)*0.65)
test_size=len(data_oil)-training_size
train_data,test_data=data_oil[0:training_size,:],data_oil[training_size:len(data_oil),:1]
```

In [38]:
```
training_size,test_size
```

Out[38]:
```
(5340, 2876)
```

In [39]:
```
train_data.shape
```

Out[39]:
```
(5340, 1)
```

creating a dataset with sliding windows

```
def create_dataset(dataset,time_step=1):
    dataX,dataY=[],[]
    for i in range(len(dataset)-time_step-1):
        a=dataset[i:(i+time_step),0]
        dataX.append(a)
        dataY.append(dataset[i+time_step,0])
    return np.array(dataX),np.array(dataY)
```

In [40]:
```
time_step=10
x_train,y_train=create_dataset(train_data,time_step)
x_test,y_test=create_dataset(test_data,time_step)
```

In [41]:
```
print(x_train.shape),print(y_train.shape)
```

```
(5329, 10)
(5329,)
```

Out[41]:
```
(None, None)
```

In [42]:
```
print(x_test.shape),print(y_test.shape)
```

```
(2865, 10)
(2865,)
```

Out[42]:
```
(None, None)
```

In [43]:
```
x_train
```

Out[43]:
```
array([[0.11335703, 0.11661484, 0.12053902, ..., 0.10980305, 0.1089886 ,
        0.11054346],
       [0.11661484, 0.12053902, 0.11550422, ..., 0.1089886 , 0.11054346,
        0.10165852],
       [0.12053902, 0.11550422, 0.1156523 , ..., 0.11054346, 0.10165852,
        0.09906708],
       ...,
       [0.36731823, 0.35176958, 0.36080261, ..., 0.36391234, 0.37042796,
        0.37042796],
       [0.35176958, 0.36080261, 0.35354657, ..., 0.37042796, 0.37042796,
        0.37879461],
       [0.36080261, 0.35354657, 0.35295424, ..., 0.37042796, 0.37879461,
        0.37916482]])
```

In [45]:
```
x_train=x_train.reshape(x_train.shape[0],x_train.shape[1],1)
x_test=x_test.reshape(x_test.shape[0],x_test.shape[1],1)
```

In [ ]: