CS353 ML Lab EndSem

Name: K V Sumanth Reddy

Roll No: 181CO225

Batch: Section 2

Date: 13/04/2021

Q: Predict Daily Deaths in Canada with the given dataset

Dataset Used: time_series_covid19_deaths_global.csv

Import Libraries

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from statsmodels.tsa.arima_model import ARIMA
#using ARIMA model - Auto Regressive Integrated Moving Average
```

```
/usr/local/lib/python3.7/dist-packages/statsmodels/tools/_testing.py:19: Futu
  import pandas.util.testing as tm
```

Data Preprocessing

```
df=pd.read_csv('time_series_covid19_deaths_global.csv')
dataset = df.loc[df['Country/Region'] == 'Canada'] #print only canad
print(dataset)
```

```
                Province/State Country/Region      Lat  ...  4/8/21  4/9/21  4
39                     Alberta         Canada  53.9333  ...    2005    2007
40           British Columbia         Canada  53.7267  ...    1493    1495
41           Diamond Princess         Canada   0.0000  ...       1       1
42             Grand Princess         Canada   0.0000  ...       0       0
43                   Manitoba         Canada  53.7609  ...     946     949
44               New Brunswick         Canada  46.5653  ...      31      32
45  Newfoundland and Labrador         Canada  53.1355  ...       6       6
46      Northwest Territories         Canada  64.8255  ...       0       0
47                Nova Scotia         Canada  44.6820  ...      66      66
48                    Nunavut         Canada  70.2998  ...       4       4
49                    Ontario         Canada  51.2538  ...    7478    7496
```

```
50       Prince Edward Island         Canada  46.5107  ...       0       0
51                     Quebec         Canada  52.9399  ...   10718   10726
52     Repatriated Travellers         Canada      NaN  ...       0       0
53               Saskatchewan         Canada  52.9399  ...     447     453
54                      Yukon         Canada  64.2823  ...       1       1

[16 rows x 449 columns]
```

```
#dropping columns and cleaning the dataset
dataset = dataset.drop(['Country/Region','Province/State','Lat','Lo
dataset.head(10)
```

	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/
39	0	0	0	0	0	0	0	0	
40	0	0	0	0	0	0	0	0	
41	0	0	0	0	0	0	0	0	
42	0	0	0	0	0	0	0	0	
43	0	0	0	0	0	0	0	0	
44	0	0	0	0	0	0	0	0	
45	0	0	0	0	0	0	0	0	
46	0	0	0	0	0	0	0	0	
47	0	0	0	0	0	0	0	0	
48	0	0	0	0	0	0	0	0	

10 rows × 445 columns

```
#adding all the canadian state death cases into one column
newdataset = pd.DataFrame(columns = dataset.columns)
newdataset.loc['Cases'] = df.sum(numeric_only=True, axis=0)
print(newdataset)
```

```
       1/22/20  1/23/20  1/24/20  ...     4/8/21     4/9/21    4/10/21
Cases     17.0     18.0     26.0  ...  2901909.0  2915400.0  2927807.0

[1 rows x 445 columns]
```

```
newdataset = newdataset.T
print(newdataset)
```

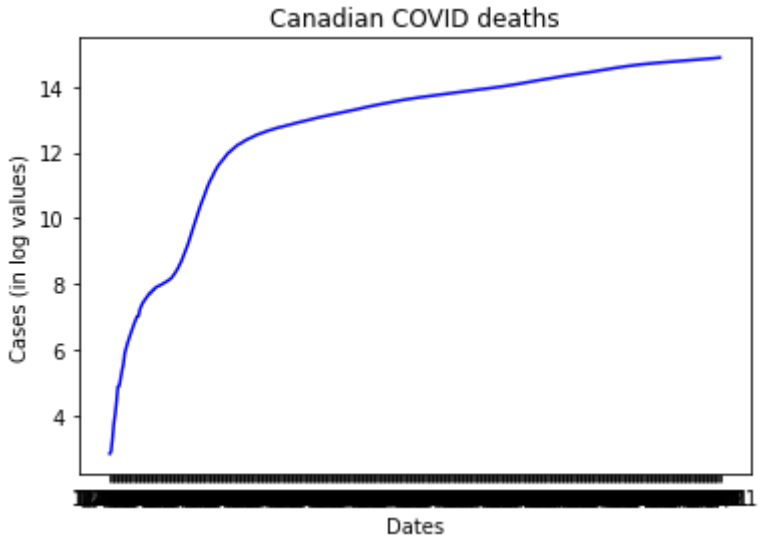
```
            Cases
1/22/20      17.0
1/23/20      18.0
1/24/20      26.0
1/25/20      42.0
1/26/20      56.0
```

```
...            ...
4/6/21   2872453.0
4/7/21   2887278.0
4/8/21   2901909.0
4/9/21   2915400.0
4/10/21  2927807.0

[445 rows x 1 columns]
```

Plotting Cases Trend

```
plt.plot(np.log(newdataset['Cases']),color= 'blue')
plt.xlabel("Dates")
plt.ylabel("Cases (in log values)")
plt.title('Canadian COVID deaths')
plt.savefig('Death Trend.png')
plt.show()
```



Training the dataset on ARIMA model and predicting cases

```
model = ARIMA(newdataset, order=(1,1,1))
model = model.fit()

#Now we predict cases 5 days ahead and 5 days behind from current da
y_hat = model.predict(len(newdataset)-5,len(newdataset)+5, typ='leve
```

```
/usr/local/lib/python3.7/dist-packages/statsmodels/tsa/base/tsa_model.py:165:
  % freq, ValueWarning)
/usr/local/lib/python3.7/dist-packages/statsmodels/tsa/base/tsa_model.py:165:
  % freq, ValueWarning)
```

Prediction values

```
#printing all the predictions of 5 days ahead and 5 days from the la
y_hat = pd.DataFrame({'Cases':y_hat})
y_hat['Cases']=y_hat['Cases'].astype('int64')
y_hat.head(10)
```

	Cases
2021-04-06	2868192
2021-04-07	2884533
2021-04-08	2901646
2021-04-09	2915547
2021-04-10	2927954
2021-04-11	2939418
2021-04-12	2950372
2021-04-13	2960753
2021-04-14	2970636
2021-04-15	2980086

```
newdataset.tail(5)
```

	Cases
4/6/21	2872453.0
4/7/21	2887278.0
4/8/21	2901909.0
4/9/21	2915400.0
4/10/21	2927807.0

Error percentage in prediction

```
temp1 = newdataset.tail(5)
temp2 = y_hat.head(5)

#Now, we print error in prediction of the last 5 day death cases wi
error = abs(temp1 - temp2)
print("Error in death cases")
print("----------------------")
print(error)
print("----------------------")
```

```
print("Error percentage")
print((error/temp1)*100)
```

```
Error in death cases
----------------------
          Cases
4/6/21   4261.0
4/7/21   2745.0
4/8/21    263.0
4/9/21    147.0
4/10/21   147.0
----------------------
Error percentage
            Cases
4/6/21   0.148340
4/7/21   0.095072
4/8/21   0.009063
4/9/21   0.005042
4/10/21  0.005021
```

We can see that the error precentage using a ARIMA model is very less and hence ARIMA model can be used to predict Canadian COVID death cases