

LSTM:

Predicting the Cases caused by covid 19 disease is a time series analysis. Covid cases are not a random process. The main issue is to forecast in such a way that there is a correlation between predicted and actual cases. LSTM is a powerful algorithm, it retains long time memory by design. The reason to choose LSTM is that the covid cases in a particular country on a particular day depend on the history of previous cases.

Explaining the steps used in the model

1. Importing Libraries:
Darts is a library used for Time Series Analysis in python, other libraries are pandas, matplotlib
2. Importing covid cases data from covid19h (Source: John Hopkins) from 1st November and relevant data for analysis:
We have used this data to merge them with the existing air traffic data prepared by the Data Engineering team.
3. Data Preparation for Analysis:
Here we merged the covid data with the cleaned data and also filtered data for certain dates.
4. Conversion to Time series:
Forecasting the covid cases is based on having a look at the history of cases and then forecasting for the future. To deal with this time series conversion into the required format is essential
5. Standardization of Data:
When the network is fed with unscaled data, the large inputs slow down the learning rate causing problems in effective learning
6. Applying Algorithms to data:
To apply algorithms, we use the dart library which takes a minimum of 10 days and forecasts 20 days in future
7. Inverse Standardization
In this step we inverse transformed the scaled data
8. Calculating error percentage $((\text{actual}-\text{forecast}) * 100 / \text{actual})$
We calculated the error rate by using above formula
9. Plotting the errors of the country as bar plots

Correlation

10. Finding Incidence Rate (no. of. cases in the country / (population of the country * time frame of consideration))
11. One hot encoding for Country names
12. Normalization of data
13. Clustering using K means clustering
14. Export the results into excel file for visualization