Deep Learning Final Project Increment 1

The dataset consists of 5 attributes:

* Province\_State – Province/State within a country or region (some as null).
* Country\_Region – Country/Region.
* Date – Date in the format of yyyy-mm-dd.
* ConfirmedCases – Integer value of confirmed cases at the date in the location of that row.
* Fatalities – Integer value of fatalities at the data in the location of that row.

As the encoding of the data goes, we have the Date attribute as a datetime object:

df\_train['Date'] = pd.to\_datetime(df\_train.Date)

Certain countries do not have any Provinces or States so for the null values of this attribute it is equated to “None”:

df\_train = df\_train.fillna({"Province\_State": "None"})

The Country/Region and the Province/State attributes are both changed to category objects:

df\_train["Country\_Region"] = df\_train["Country\_Region"].astype('category')

df\_train["Province\_State"] = df\_train["Province\_State"].astype('category')

The model that we would use would be simple, an RNN. RNN as the choice because we are predicting future data based on patterns. The model as is would be extremely inaccurate because of the little data given in the initial dataset so data must be added. Data we could add to the training dataset include density of the area, 1st or 3rd world , when the country goes on lockdown, hospital beds per capita, and type of government, etc.