# 



SCHOOL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCE (SEECS)

**Machine Learning**

|  |
| --- |
| **Project Report**  **Forecasting Covid-19** |

***Submitted by:***

***Usman Khawar 208775***

***Submitted to:***

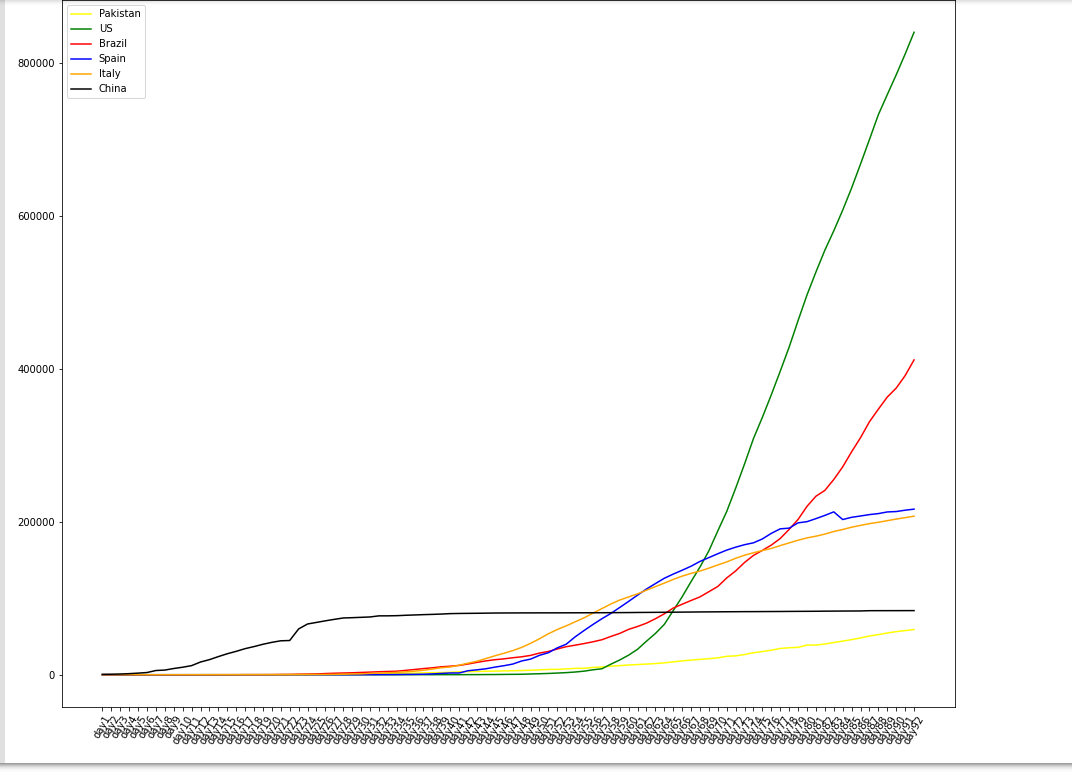
Sir Faisal Shafait

**Class**: BESE-8B

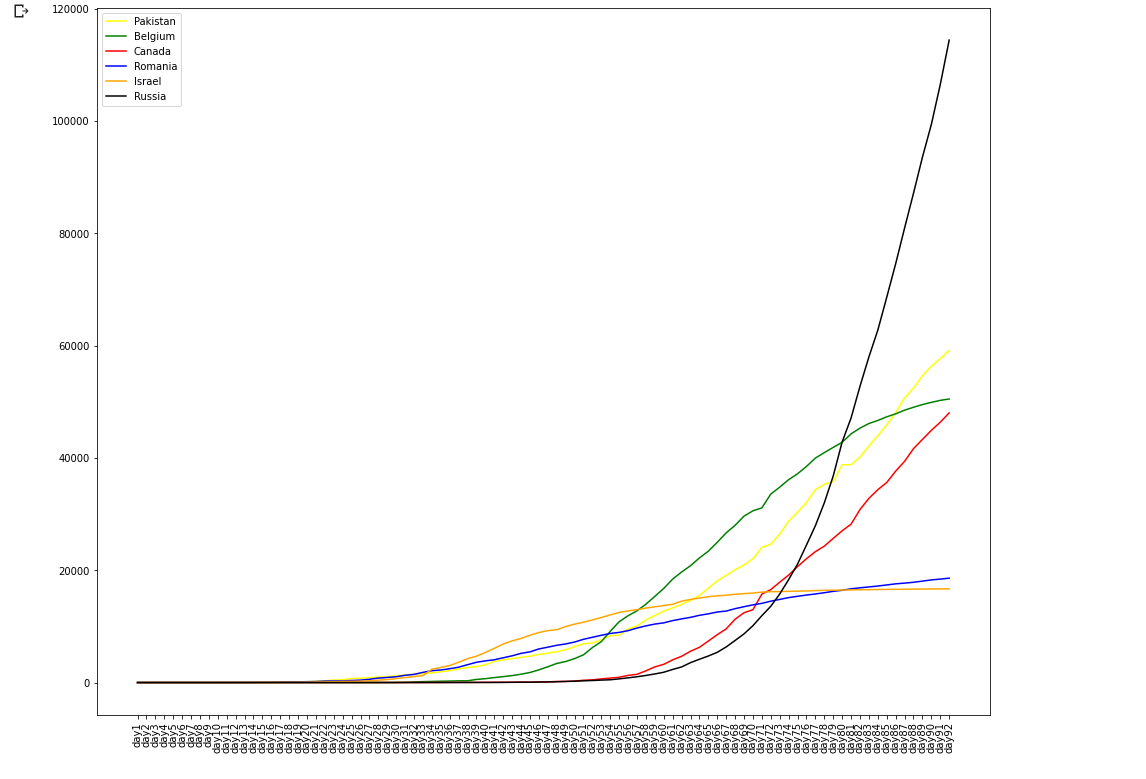
**Dated:** 06-11-2020

# Approach:

After importing the data frame, and adjusting it according, as according to the requirement we are supposed to find the *number of days versus amount of cases as a measure of spread. So, I framed it according to Pakistan by removing the trailing zeros which comes out to be 92 days, then first I used correlation technique to find the related countries, which turns out to be not so suitable when plotted*



Then I used the approach of absolute differences comparing country time series with Pakistan using NumPy array, the countries extracted by this logic tends to be more suitable for comparison when plotted:



We save a new data frame for the valid countries of comparable length of Pakistan. Now it is time to model the time series, our time is **commutative**, so in order to convert it to **individual entities** we take corresponding differences, Now we normalize the data and then use RNN for the model, I could have used LSTM but turns out I lack expertise to model it, I have to take peer help for the modelling functions, the logic that I was struck was how to predict 30 days in number, turns out we use a loop and train a value for each. Where we predict new cases for 30 days considering the previous outputs.

10th May until 27th May is used as validation set and the graphs are plotted

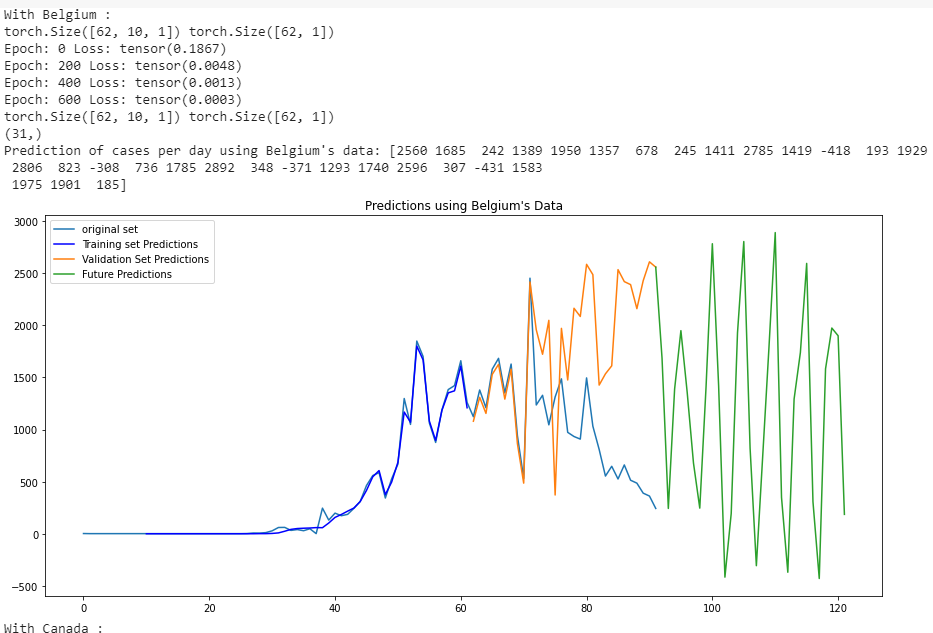
# Result:

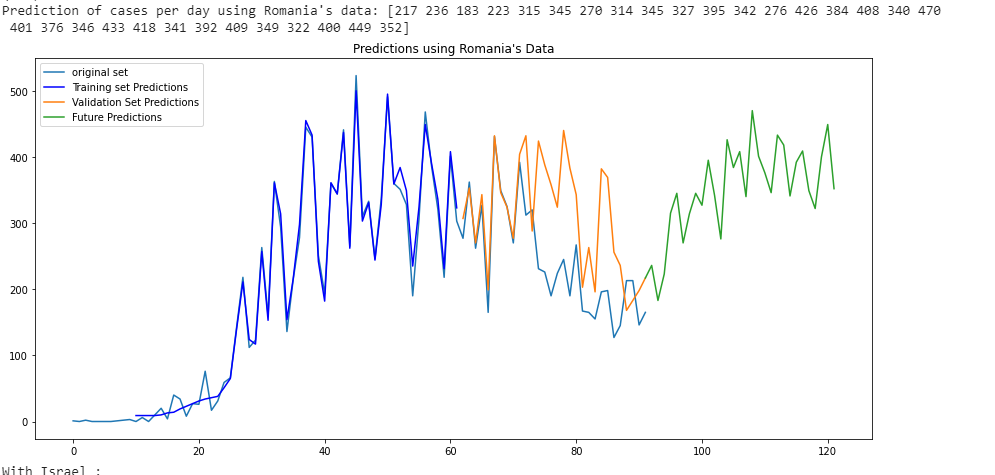
I have to do a lot of hit and trail for parameter tuning, using the Adams optimizers and

Finding out the following parameters are best suited

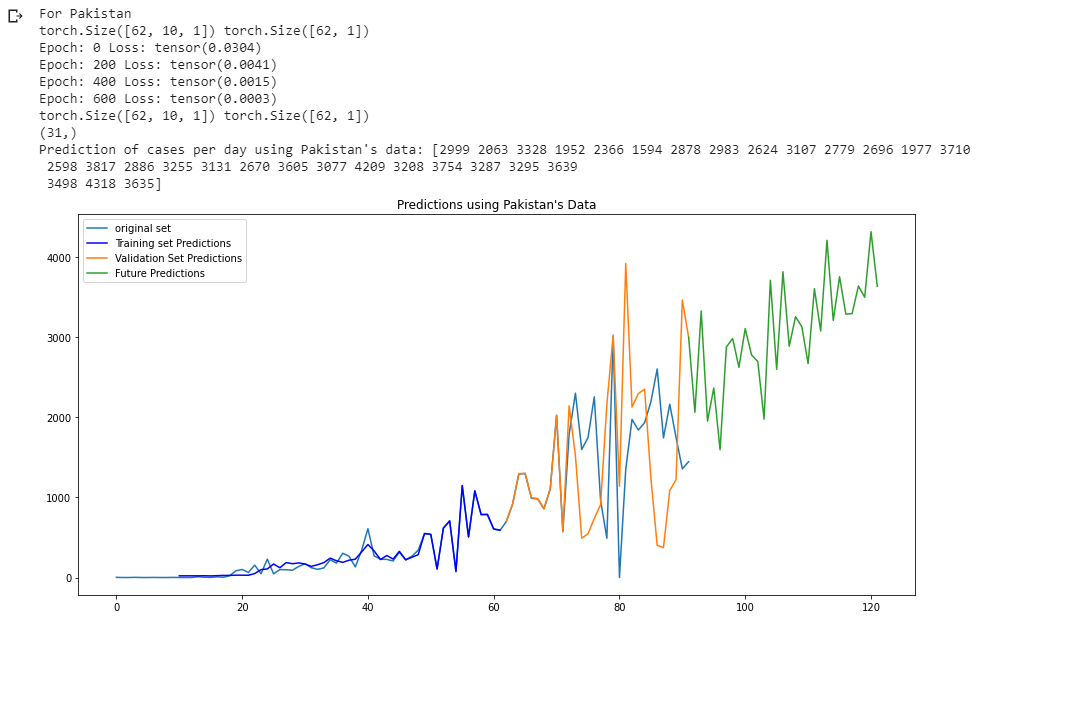
one input, one output and 3 hidden layers with 90 neurons each. Model is trained with 0.002 learning rate for 600 epochs. mean squared error is used as parameter for error. The sequence length fed is 10.

It nearly took half a day for tuning right combination which was kind of frustrating.





Out of five countries, I have attached snippets of two, as clearly seen the model is not over fitting as validated from validation set predictions but the model is **fluctuating** and tends to replicate the **previous pattern** with higher **spikes** (variations).



At each hit and trail the results are different by a comparable amount, my model does **predict -vet cases**

**Wh**ich is sort of a bug and I still don’t have an idea why it happens, but I will definitely try to look into this problem.

# Findings

In Particular to Pakistan, the results aren’t handy, There are up and down spikes which is not the case in comparison to what we are actually seeing, further more one cannot predict model based on other countries data, because the dataset doesn’t take into account the Lockdown period and situation as well as the Sops being followed, so it isn’t a reliable parameter. Furthermore, prediction of negative values is sort of a bug, which I will consult later just to make sure it is not termed as plagiarism. The dataset is missing vital information too hence the reliability is not as par desired.

Note everything is due credited in the ipynb file, please take into account before giving straight zeros as before. I have put in a considerable effort for all these three assigments