

**Machine Learning**

**Project Report**



**Name:** Saad Akram

**Class:** BESE-8A

**CMS:** 221434

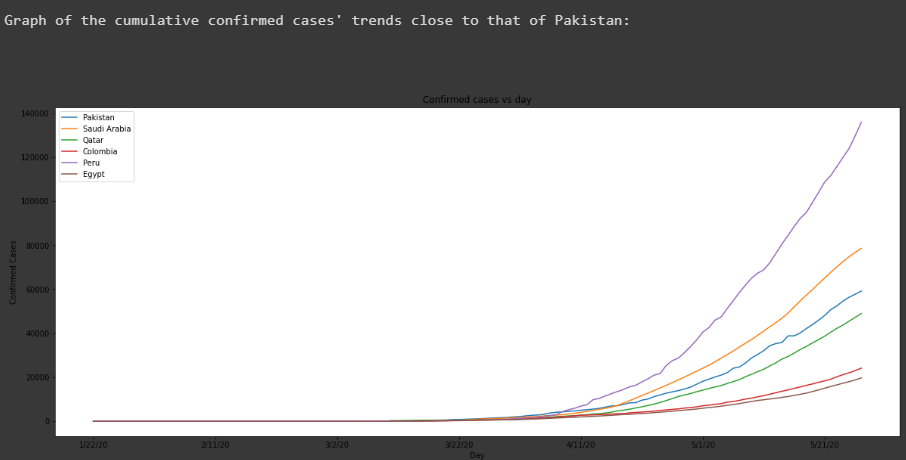
**Project Description and approach**

The project we were assigned was to examine the dataset having the confirmed cases of COVID-19 and predict the future trend of Pakistan according to it. We had to use any model from amongst those that were being given in the document. So, the basic task was to extract the top five countries having the same trend as that of Pakistan. The countries after extraction were; Saudi Arabia, Qatar, Colombia, Peru, Egypt. And then we had to use the data of these countries and had to predict the cases of Pakistan from 10th May till 27th June.

All the task given in the project description are explained as follows according to the approach I used to implement the LSTM model and then, predicting the values. The approach has been discussed under the heading of each task

**Task 1,2**

In this task we had to analyze the entire dataset and separate out the countries having the same trend as that of Pakistan. What I have done is that I grouped the entire dataset according to the countries and then taken the topmost countries having the same trend as that of Pakistan and the trend of these countries can be seen by the following graph. As, the data was cumulative hence, this graph is cumulative trend.

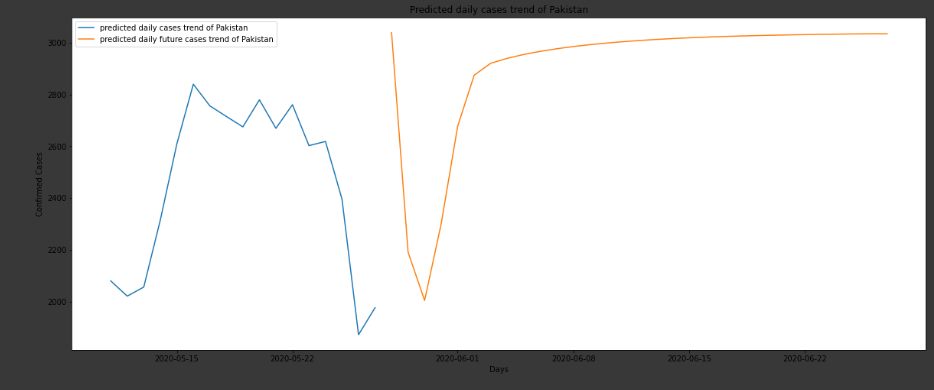


**Task 3**

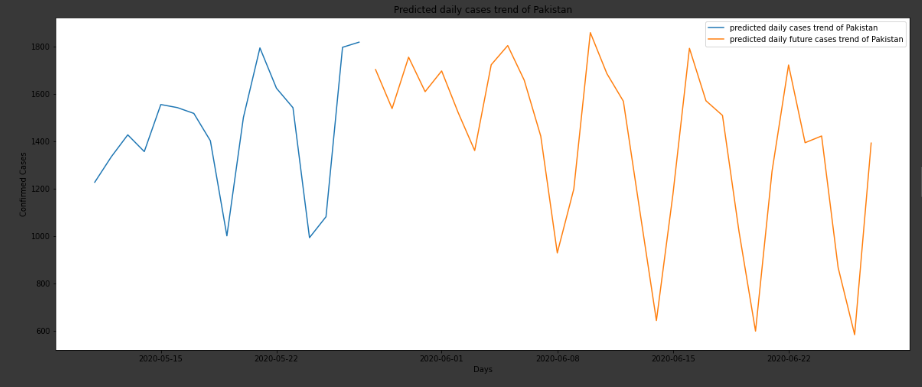
In this part we had to select a model train our model and predict the cases of Pakistan. Hence, the model I selected is **LSTM.** I trained the model over all the dataset of the countries separately and then predicted for Pakistan from 27th May till 31st June according to the trained model for every country getting separate predictions for all. For forecasting into future, the approach, I developed was that I selected the dataset for previous 31 days and then looped it over number of days which is 31 in my case. At each iteration I entered the data in my already trained model, got a prediction according to the. Then, I added that prediction in my dataset of previous cases and moved the window one step forward. Then, the latest 31 prediction were my days in future.

Following are the graphs of Pakistan as predicted by the other countries;

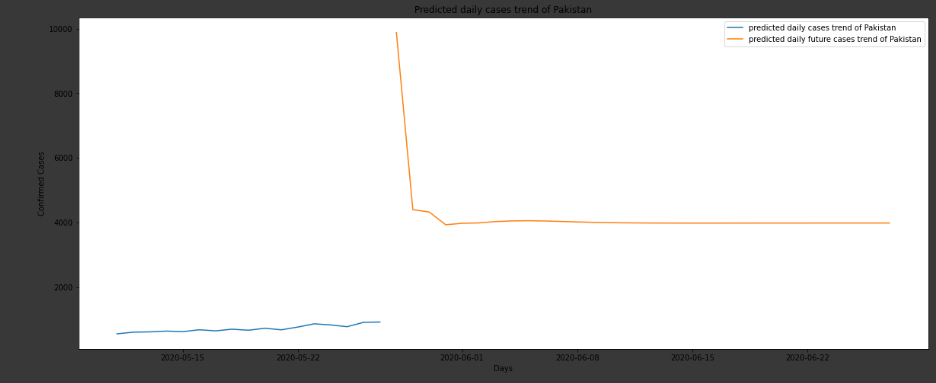
1. Saudi Arabia



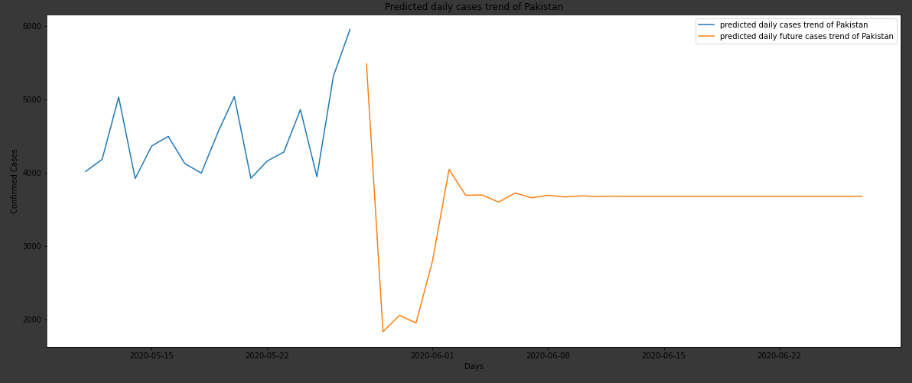
1. Qatar



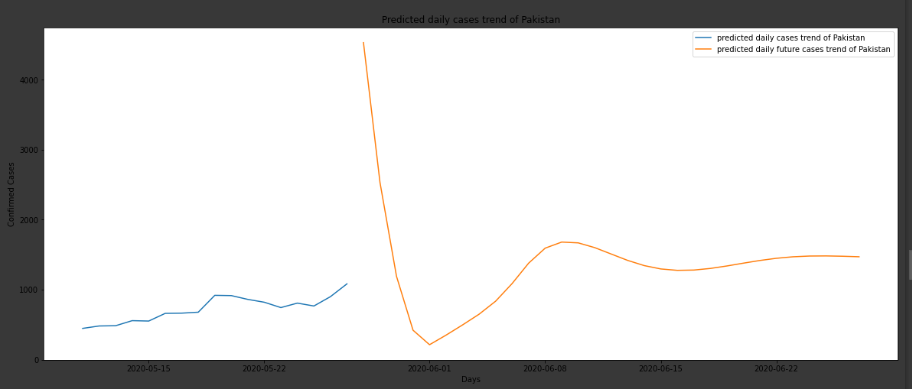
1. Colombia



1. Peru



1. Egypt



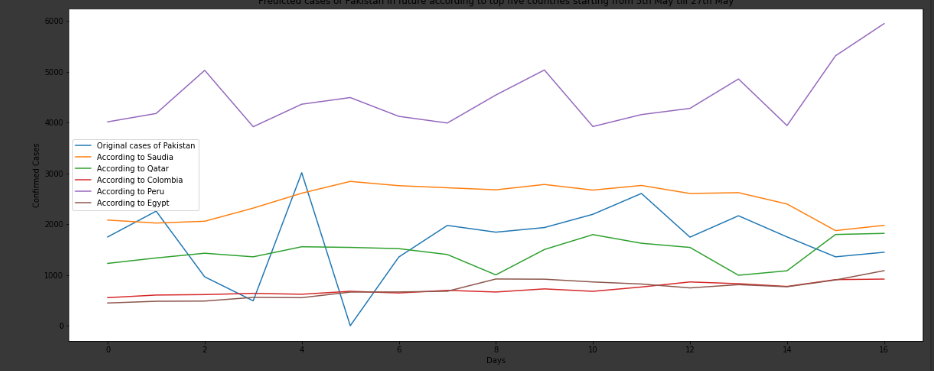
In all the above graphs, the blue line shows the predicted cases from **10th May** till **27th May**. And the orange line shows the predicted cases from **27th May** till **27th June**. Apart from **Saudi** all other countries have predicted an increase in the daily cases from **10th May** till **27th May**. And all countries have predicted an approximate decrease in the cases from **27th May** till **27th June** for Pakistan. This decrease is fluctuating, and this shows that the model has high capacity. The list of the predicted cases from **10th May** till **27th June** for Pakistan can be seen in the output under the heading of **Part 4 ( .ipynb file).**

**Task 5**

For, this task we need to see that how well the model has performed which can be seen by the following graph. And in order to examine it we plotted the predicted and the original graph and the result be follows.

I calculated the mean square error (MSE) for the predicted values from the original daily values from **10th May** till **27th May.** And the MSE shows that the predictions made by Qatar are close to that of Pakistan and has performed well.

**All Graphs from 10th May till 27th May**



Hence, the model has performed good like it had been able to spot the rise and the fall in the number of daily cases according to the number of days. The **Blue** line in the above graph shows the original cases of Pakistan as given in the dataset. And the graph shows that the daily cases are fluctuating and hence, so is the graph.

The graph shows that the **Green** line by Qatar has been able to somehow predict the trend of Pakistan. Like, at intervals it has predicted the exact rise and the fall in the number of cases and overall it has shown the increase in the cases as that can be seen in the original dataset of Pakistan.

Colombia and Egypt having **red** and **Purple** lines respectively have not performed well and somehow could not been able to predict well for Pakistan. Overall, they have shown increase in the overall daily cases just like Pakistan.

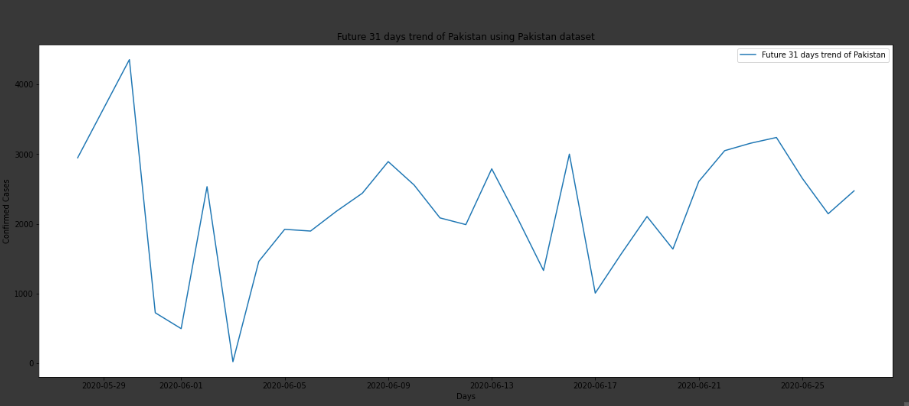
The **Yellow** line of Saudi Arabia has somehow tried to predict according to the original cases of Pakistan. At intervals, it had been able to capture the rise and fall of the cases. Overall, the predictions show a minimal decrease in the overall cases.

The line for **Peru**  has followed somewhat the similar trend of rise and fall according to that of Pakistan but, the difference of predicted cases are miles apart from the original cases, but the model has predicted somewhat similar trend for Peru.

The model is not 100% reliable and the reason can be due to the minimal dataset we have and the unpredictable daily number of cases because a lot could not have been researched about the activity of the virus. But the model has been able to capture the trend of rise and fall of daily cases at multiple intervals.

**Task 6**

For this task we had to train the model on the dataset of Pakistan and make predictions. For this task I made prediction from **27th May** till **27th June**. The procedure was the same. The following graph shows the predict trend 31 days into the future.

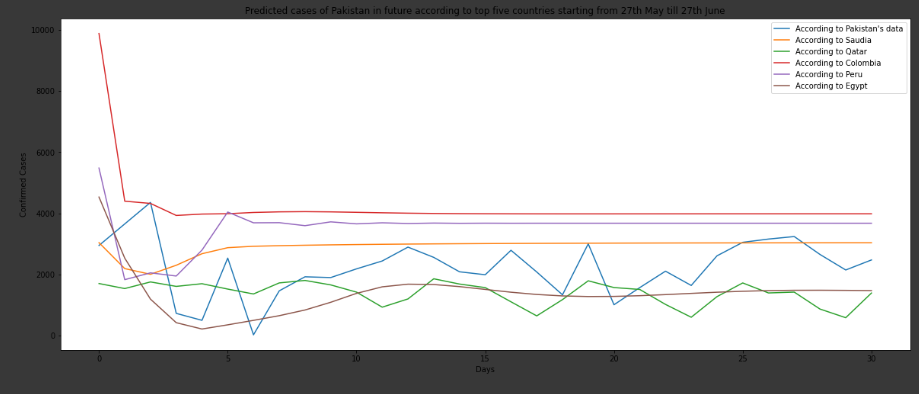


The graph shows fluctuations in the number of daily cases but, this shows that the overall cases would decrease in future from almost **3000** to **2100** daily cases but, it shows that the daily cases would be fluctuating in the future according to the previous daily cases of Pakistan. The last few days of June shows that the cases would start to decrease a little.

**Task 7**

For this task I plotted all the graphs the predictions of Pakistan from **27th May** till **27th June** (31 days in future) according to the top five countries having the same trend as that of Pakistan and the predicted trend according to the data of Pakistan. Following is the graph and the explanations.

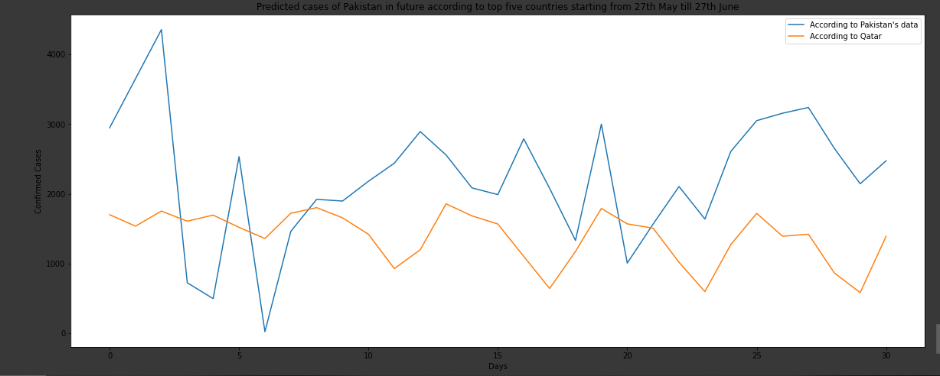
**All Graphs from 27th May till 27th June**



The **Blue** line shows the trend of Pakistan according to the data of Pakistan and we can see the graph is fluctuating but the we can see the overall decrease in the daily cases of Pakistan. It can be seen from above that all the countries having the colored line except green shows that after almost 15 days the number of cases is almost constant. These countries include Saudi, Peru, Colombia, Egypt. But all the countries show that the cases they predicted will tend to decrease with rise and fall at intervals but till 30th day the overall daily cases tend to decrease. The line of **Egypt** the **purple** line has somehow tried to copy the trend of Pakistan but after a while becomes constant with little fluctuations. The **Yellow** line of **Saudi Arabia** has predicted almost approximately the same number of daily cases as that of the original of Pakistan, but the trend of rise and fall isn’t followed

**Best Predictions**

According to the graph we can see that the **green** line that of **Qatar** has been able to predict the trend like that of Pakistan from the original dataset. **Qatar** has been able to capture the rise and the fall in the daily cases like that of the original case of Pakistan. And the trend is almost the same hence, the model has predicted well. Hence, **Qatar** has predicted the best for the future cases of Pakistan. It has almost predicted the same trend in the last **18 days** from **12** to **30** days period. This can be seen in the graph above. Following is the separate graph for **Qatar** predicted days and the days predicted by **Pakistan.**



**Results and Findings**

The model is not 100% reliable because the dataset was minimal only approximately the length of 127. But it has been able to capture the trend which is considered somehow a success. And according to it. The result is that the dataset of Qatar has been able to make the predictions according to the trend of Pakistan and is that closest and it shows that the daily cases would show fluctuations. The situation would not be better in one-month future like the cases would be fluctuating and there are many reasons for that, and this might include the reasons for the lifting of Lockdown and people going out for shopping on Eid. But the overall daily cases would tend to decrease and there will be less at the end of month of June then they were at the start of June. But we need to keep ourselves and the people around us safe as this is a deadly virus, it is real and killing people daily.

**Conclusion**

This project was a great learning process and it helped me to get a hand over the machine learning algorithms and the proper implementation of the models. Before this course I would see machine learning as fancy but, hard word. But, now after the projects, labs, assignments it has somehow eliminated the fear I had previously and helped me see how the daily and future predictions are made. And I am grateful to learn out of this course as the had expected before starting it.