Data Visualization of Covid-19 & Comparative Study on South Asian Countries

BY

Sifat Ul Haque
ID: 183-15-2289

AND

Toufiq Islam Swoad
ID: 183-15-2282

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Dr. S.M. Aminul Haque
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised By

Mushfiqur Rahman
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
SEPTEMBER 2022

APPROVAL

This Project titled "**Data Visualization of Covid-19 and Comparative Study on South Asian Countries**", submitted by **Sifat Ul Haque, Id: 183-15-2289** and **Toufiq Islam Swoad, Id: 183-15-2282** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 13th September 2022

BOARD OF EXAMINERS

Dr. S M Aminul Haque — **Chairman**
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

13/09/22

Fahad Faisal — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Fourcan Karim Mazumder — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Mohammad Shorif Uddin — **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. S.M. Aminul Haque,** Associate Professor, Department of CSE, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Dr. S.M. Aminul Haque

Associate Professor and Associate head

Department of CSE

Daffodil International University

Co-Supervised by: 16.09.22

Mushfiqur Rahman

Lecturer (Senior Scale)

Department of CSE

Daffodil International University

Submitted by:

Sifat Ul Haque

ID: 183-15-2289

Department of CSE

Daffodil International University

Toufiq Islam Swoad

ID: 183-15-2282

Department of CSE

Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible for us to complete the final year project/internship successfully.

We are really grateful and wish our profound indebtedness to our supervisor **Dr. S.M. Aminul Haque**, Associate Professor, Department of CSE Daffodil International University, Dhaka. Deep Knowledge and keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Mushfiqur Rahman,** Lecturer (Senior Scale)**,** Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

The Coronavirus had disrupted our day-to-day lives when it was first reported back in 31 December 2019. Lifestyle to economy, everything tumbled for most countries. It has been a long time since then and countries have taken measures to combat it and now almost all countries are restoring their economy and people's lifestyles that were hampered due to the lockdowns. Developed countries with their better medical facilities did struggle to control the outbreak of the virus. In South Asia, where countries, being mostly under-developed or developing, took some massive hits from the pandemic that lasted for more than 2 years. In this study, we do a comparative study on the South Asian countries to find out what were their pandemic conditions in terms of confirmed cases, recovered cases and death cases. Also analyze which countries took appropriate measures to do comparatively better to handle the pandemic from the rest. Time series and prediction model will be used to do the comparative analysis. We used a 4-months phase analysis along with other plots and graphs to analyze the COVID-19 spread for the countries and for predicting the curve we used TensorFlow. The timeframe we chose to analyze is between 23-01-2020 to 08-03-2021. Developing countries can get the idea from this article how to tackle a pandemic or epidemic in the future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

The Coronavirus first affected us in December 2019. The spread of the virus started from the Wuhan city of China. From there point it started to spread rapidly all over the world. A lot of lives were lost within a few months. The rate of contagion and spread of infection is quite fast compared to other viral infections encountered until today. Due to its rapid progress and covering the world in a short period of time, it is necessary to carry out intensive studies on it to understand it properly. The data used includes 8 countries of South Asia (Afghanistan, Bangladesh, Bhutan, India, Nepal, Pakistan and Sri Lanka) where the disease progresses differently. We have visualized the confirmed, death and recovered cases of these countries for a comparative study along with a prediction model to understand the future possibilities of the virus spread.

1.2 Motivation

We all had been suffering from the miseries of Covid-19 for more than 2 years now. This pandemic has definitely taught us the necessity to study and research on relevant topics like Covid-19 and similar potential pandemics that might arise in the future. It also motivated us to look into the data of Covid-19, analyze it deeply and get some insights for knowing the risk factors and spread of it in a developing country like Bangladesh and so on in South Asia. The adverse effects of it and the critical situations that nobody was prepared for made us also concerned to have a proper analysis of such pandemic so that we can tackle similar ones better in the future. This is our core motivation to work on such research-based project as we find it relevant and appropriate for the time being.

1.3 Rationale of the Study

Everyone is trying to understand COVID-19 and why it caused so much damage to humanity in this century. There are researches going around the globe on it right now. Our motive is to do a proper research and analysis of the South Asian countries so that we can

get a better understanding of the situation in our regions. As developing countries, we need to be prepared for the future if such a pandemic ever repeats itself, we are better prepared to tackle the situation. We can learn from our mistakes and what our neighbors did right to reduce the damage caused by the virus in their respective countries.

1.4 Research Questions

- Could the situation be handled better when the virus started spreading?
- Were the countries prepared to face such a pandemic?
- Was the medical infrastructure well-built for providing treatment to people?
- What other countries did comparatively better and reduced their damage caused due to the pandemic?
- How can we be better prepared for such a pandemic in the future?

1.5 Research Objectives

- Non-misleading information and provide clarity on the situation.
- Provide accurate context and depth of collected data.
- Prediction of future curve of death/infection/healed from present and past data.
- Visualization of a pandemic like COVID-19 will help to understand how the pandemic spread, rate of spreading and how different countries handled the situation might help in future to take proper precaution to reduce potential pandemic damage in the future.

1.6 Expected Outcome

Visual representation of rate of death, rate of infected people and recovery rate.

Comparative analysis on the basis of these figures will enable us to monitor and understand the outbreak of any similar pandemic in the future, with appropriate precautions and steps to reduce loss of lives as well as financial downfall.

1.7 Report Layout Chapter

The paper is organized as follows:

I. Background
II. Research Methodology
III. Experimental Results and Discussion
IV. Summary, conclusion, Recommendation and implication for future Research
V. Reference.

CHAPTER 2

Background Study

2.1 Terminologies

Our main goal of this paper is to understand the COVID-19 for South Asian countries by using data visualization tools like scatter plot, bar plot, line plot, choropleth map, tree map and so on. Later, we used ARIMA model and Facebook Prophet to do prediction for upcoming days how the pandemic will affect in terms of confirmed cases, death cases and recovered cases. Another thing we implemented here is a 4-month phase plot that will visualize the scene on a quarterly basis for the countries. We also did per 100 cases studies to get a better understanding. On the later part, we did some comparative study based on decisions taken by some countries which gave advantages and disadvantages over other countries for minimizing the pandemic effect on them or failed to do that. We took inspiration from some notable researchers who are doing tremendous work to understand COVID-19 from their respective countries. Our vision is the same to tackle such pandemic better in the future.

2.2 Related Works

Machine learning and Artificial Intelligence (AI) models are essentially used to improve the prediction accuracy of diagnosis and the screening of non-infectious diseases. Moreover, machine learning approaches are also widely used in the evaluation and prediction of COVID-19 survival rate, and the release time of patients based on clinical data.

Malki, Z., Atlam, ES., Ewis have advanced their studies on a decision tree algorithm on the COVID-19 global actual-time data. The core idea is to utilize supervised machine learning algorithms for time-series forecasting. The algorithms proposed in this work, namely: the decision tree algorithm and linear regression, are powerful models for predicting sequence and time-series data-related problems. [1]

Sahai, A.K.Rath, N.Sood, V., and Singh used the ARIMA model where specifications had been envisioned using the Hannan and Rissanen algorithm. The sample forecast for the following 77 days was computed using the ARIMA models. [2]

Gopi Battineni, Nalini Chintalapudi, Francesco Amenta applied Facebook Prophet for predicting Covid-19 data since this model perfectly fits for historical data of several seasons data and strong seasonal effects and it is fully automatic with limited manual involvement. A well-derived Prophet model not only helps to future predictions but also to detect anomalies and fill gaps in missing values. [27]

Sun, J., Chen, X., Zhang, Z., Lai, S., Zhao, B., Liu, H., Wang, S., Huan, W., Zhao, R., Ng, M.T.A. and Zheng, Y proposed a new model named Dynamic Susceptible Exposed Infective Quarantined (D-SEIQ), by making appropriate modifications of the Susceptible-Exposed-Infected Recovered (SEIR) model and integrating machine learning-based parameter optimization under epidemiological rational constraints to predict the long-term reported cumulative numbers of COVID-19 cases in China. [3]

Shaito, M. and Elmasri, R. focus on spatial data visualization in their research. They used six techniques to visualize spatial data namely: Choropleth maps, Heat maps, Hexagonal binning, Dot maps, Bubble maps, and Cartogram maps. [5]

Ogundokun, R.O. and Awotunde, J.B. projected a simple average aggregated scheme, and the aggregated system has been established by aggregating three regression methods which include Support Vector Regression, and Linear Regression as well as Artificial Neural Network. The variables utilized for the formulation of an aggregated method are the figures of COVID-19 cases. [6]

After everything considered, we proposed that the best way to understand the spread of COVID-19 is by using ARIMA & Facebook Prophet models to predict the spread of the virus. [7]

2.3 Research Summary

Pandemic can cause massive damage to any country. And if the country is a developing or underdeveloped country, then the damage done will be more adverse compared to a developed country.

Our paper focuses on the South Asian countries, where mostly we see developing nations with limited resources in medical sectors. So pandemics can cause some severe damage to them. And for the case of COVID-19 it was not different, we saw reported deaths almost

daily from these countries when it started to spread rapidly. Bangladesh, India and Pakistan suffered considerably more compared to other South Asian countries when we see the death cases per 100 cases.

The lockdown calls from South Asian countries helped them initially to reduce the spread of the virus. But still a lot of damage was done to the economy and there was a significant loss of lives. People couldn't get proper health facilities due to infrastructural limitations in countries like Bangladesh, Pakistan and India due to dense population. Other South Asian countries comparatively faced lower numbers of deaths and economic downfall due to proper measures from their respective governments in time. Bhutan and Maldives did significantly better in this regard. When we saw the cases for Sri Lanka, Nepal and Afghanistan they had lower deaths than Bangladesh, India and Pakistan as their population was not that densely populated and their medical sector could somewhat cope up with the situation. We definitely saw reports on lack of ICUs and Oxygen cylinders along with medical supplies across the countries.

Therefore, we can say there is a need for proper infrastructural developments in the medical sector for South Asian countries to face such pandemic in the future. Doctors per capita is also something we saw adequate deficiency for most countries. [10]

Apart from these, our comparative study shows policymakers could further work on their current medical sector policies to ensure proper facilities for their citizens. If proper measures are taken and changes are done by the countries, then in future they can tackle such pandemics much better with minimum loss of lives and economic downfalls.

2.4 Scope and Challenges

A lot of research has been done and are ongoing on COVID-19. This is our attempt to understand the effects of it on South Asian countries on the basis of a comparative study. We analyzed the data that we could source from 23-01-2020 to 08-03-2021, along with relevant papers that focused on the study of the virus in terms of spread and damage to these countries.

Recently we have seen that COVID-19 is mutating itself and variants are coming out, Omicron and Delta are the two notable ones among them. Our study couldn't take the

variants under consideration. And these variants caused change in death rates and affection rates. The mortality rate has been going down recently with increased confirmed cases. There is also a significant number of people who are recovering from COVID-19 faster due to vaccination and antivirals being readily available to people, which wasn't the case when the virus first started its spread. This is also reflected in our prediction model where we see an increased number of confirmed cases but death cases along with steady recovered cases.

CHAPTER 3

Research Methodology

3.1 Introduction

Our goal of the paper was to do a comparative study based on time series and prediction model to understand the spread of COVID-19 in South Asian countries. For this we used various plots and graphs that are available in python including Plotly, Matplotlib, Folium and Seaborn. All these were open-source, so we could use them without any problem.

We used a bar diagram model to do a per 100 cases study for the countries to understand the spread of the virus based on 4 months' time intervals. Choropleth map was used to see the change visually over time. Other visualizations were also done to understand the changes among the countries.

For the prediction model, we used the ARIMA & Facebook Prophet to do a prediction for confirmed, death and recovered cases of total South Asian countries.

From these studies, we were able to come to an understanding about the spread of the virus over time.

3.2 Data source

The data used in this study were preprocessed and collected from open GitHub repository, and it is sourced from this upstream repository maintained by the team at Johns Hopkins University Center for Systems Science and Engineering. The upstream dataset currently lists the following upstream data sources: World Health Organization (WHO), BNO News, etc. This data shows the daily total COVID-19 confirmed positive cases, daily and total deaths, and the total and daily recoveries.

3.3 Data Collection Procedure

Data we used were collected from Johns Hopkins University Center for Systems Science and Engineering GitHub repository. They compiled their data from various sources and pre-processed for research purposes. So, we directly collected it from there.

3.4 Data Preparation

The data that we have collected was containing data for all countries in the world. Since we are working in only eight countries of South Asia, we just kept those 8 countries for our working purpose. Then this data was divided into four phases (by 4 months approximately) respectively for showing the change of COVID-19 situation of the total population of South Asian countries.

Shape of Data (473, 1)

Date	Confirmed
2020-01-22	0
2020-01-23	0
2020-01-24	0
2020-01-25	0
2020-01-26	0
...	...
2020-04-26	5416
2020-04-27	5913
2020-04-28	6462
2020-04-29	7103
2020-04-30	7667

Figure 3.1: Sample Dataset

```
ARIMA(5,2,1)(0,0,0)[0]             : AIC=6759.315, Time=0.21 sec
ARIMA(5,2,3)(0,0,0)[0]             : AIC=6697.821, Time=1.79 sec
ARIMA(5,2,4)(0,0,0)[0]             : AIC=6666.304, Time=1.87 sec
ARIMA(4,2,4)(0,0,0)[0]             : AIC=6701.063, Time=1.71 sec
ARIMA(5,2,5)(0,0,0)[0]             : AIC=6715.655, Time=1.87 sec
ARIMA(4,2,5)(0,0,0)[0]             : AIC=6686.846, Time=1.97 sec
ARIMA(5,2,4)(0,0,0)[0] intercept   : AIC=6667.831, Time=2.31 sec

Best model:  ARIMA(5,2,4)(0,0,0)[0]
Total fit time: 19.552 seconds
```

SARIMAX Results

Dep. Variable:	y	No. Observations:	473
Model:	SARIMAX(5, 2, 4)	Log Likelihood	-3323.152
Date:	Mon, 15 Aug 2022	AIC	6666.304
Time:	13:48:15	BIC	6707.852
Sample:	0 - 473	HQIC	6682.649

Covariance Type: opg

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.4337	0.055	7.886	0.000	0.326	0.542
ar.L2	-1.1053	0.061	-18.221	0.000	-1.224	-0.986
ar.L3	0.2316	0.078	2.952	0.003	0.078	0.385
ar.L4	-0.6841	0.047	-14.531	0.000	-0.776	-0.592
ar.L5	-0.3409	0.035	-9.610	0.000	-0.410	-0.271
ma.L1	-0.6965	0.059	-11.774	0.000	-0.812	-0.581
ma.L2	1.2084	0.075	16.060	0.000	1.061	1.356
ma.L3	-0.5727	0.081	-7.085	0.000	-0.731	-0.414
ma.L4	0.8185	0.058	14.227	0.000	0.706	0.931
sigma2	9.545e+04	3744.240	25.493	0.000	8.81e+04	1.03e+05

Ljung-Box (L1) (Q):	0.08	Jarque-Bera (JB):	1232.59
Prob(Q):	0.78	Prob(JB):	0.00
Heteroskedasticity (H):	3.69	Skew:	-0.42
Prob(H) (two-sided):	0.00	Kurtosis:	10.88

SARIMAX Results

Dep. Variable:	value	No. Observations:	443
Model:	ARIMA(5, 2, 4)	Log Likelihood	-3044.953
Date:	Mon, 15 Aug 2022	AIC	6109.907
Time:	13:48:17	BIC	6150.797
Sample:	01-22-2020 - 04-08-2021	HQIC	6126.037

Covariance Type: opg

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	2.0619	0.120	17.213	0.000	1.827	2.297
ar.L2	-2.0240	0.229	-8.848	0.000	-2.472	-1.576
ar.L3	1.0208	0.188	5.422	0.000	0.652	1.390
ar.L4	-0.2390	0.095	-2.510	0.012	-0.426	-0.052
ar.L5	0.1776	0.066	2.692	0.007	0.048	0.307
ma.L1	-2.4431	0.135	-18.142	0.000	-2.707	-2.179
ma.L2	2.7461	0.304	9.031	0.000	2.150	3.342
ma.L3	-1.5786	0.298	-5.289	0.000	-2.164	-0.994
ma.L4	0.3238	0.128	2.529	0.011	0.073	0.575
sigma2	6.652e+04	2808.009	23.689	0.000	6.1e+04	7.2e+04

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	771.59
Prob(Q):	0.94	Prob(JB):	0.00
Heteroskedasticity (H):	2.10	Skew:	-0.37
Prob(H) (two-sided):	0.00	Kurtosis:	9.44

Figure 3.2: Building and training ARIMA model

	ds	trend	yhat_lower	yhat_upper	trend_lower	trend_upper	additive_terms	additive_terms_lower	additive_terms_upper	weekly	weekly_lower	weekly_upper
0	2020-01-22	-2182.282051	-31161.051525	29054.178476	-2182.282051	-2182.282051	93.963731	93.963731	93.963731	93.963731	93.963731	93.963731
1	2020-01-23	-2128.066275	-36327.498949	27367.103149	-2128.066275	-2128.066275	224.495598	224.495598	224.495598	224.495598	224.495598	224.495598
2	2020-01-24	-2073.850499	-32139.591052	27774.588843	-2073.850499	-2073.850499	309.243331	309.243331	309.243331	309.243331	309.243331	309.243331
3	2020-01-25	-2019.634724	-33191.615746	27077.958727	-2019.634724	-2019.634724	142.186338	142.186338	142.186338	142.186338	142.186338	142.186338
4	2020-01-26	-1965.418948	-32433.687115	28142.118234	-1965.418948	-1965.418948	-419.478014	-419.478014	-419.478014	-419.478014	-419.478014	-419.478014
...	...	...	...	...	...	...	...	...	...	...	...	...
498	2021-06-03	804860.013221	769736.227835	843528.102856	784993.567714	826679.184293	224.495598	224.495598	224.495598	224.495598	224.495598	224.495598
499	2021-06-04	807318.046441	774489.005123	846168.485446	786431.780242	830049.291008	309.243331	309.243331	309.243331	309.243331	309.243331	309.243331
500	2021-06-05	809776.079661	773240.849867	847535.356242	787842.038035	833923.544870	142.186338	142.186338	142.186338	142.186338	142.186338	142.186338
501	2021-06-06	812234.112881	776276.748942	848714.647373	789244.956953	837707.206953	-419.478014	-419.478014	-419.478014	-419.478014	-419.478014	-419.478014
502	2021-06-07	814692.146101	776927.025001	853361.649499	790618.114032	842100.321407	-258.143943	-258.143943	-258.143943	-258.143943	-258.143943	-258.143943

Figure 3.3: Building Facebook Prophet model

3.5 Research Subject and Instrumentation

Our research topic is **Data Visualization of Covid-19 and Comparative Study on South Asian Countries,** which is mainly focused on data visualization from the data of COVID-19 of South Asian countries. In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions. Python is a very popular language for data manipulation related stuffs. With its vast library that's open-source, mostly helps researchers to get access to them & use them for research related purposes. Also, we get flexibility using python because it can handle a lot of things by itself. Thus, we have used python for doing the visualization for our research.

We have also predicted from those data using python with the help ARIMA model. An ARIMA model is a class of statistical models for analyzing and forecasting time series data. It is an acronym which stands for Autoregressive Integrated Moving Average. It is the generalization of the simpler autoregressive moving average, and that's the notion of integration. An ARIMA model is characterized by three parameters: p, d, q where,

- p is the order of AR, which forecast a series based on the past values in the series. Simply it is the number of lag observations in the model which is also known as the lag order.

- q is the order of MA, which forecast a series based on the past errors in the series. Simply it is the size of the moving average window which is also known as the order of the moving average.
- d is the number of differencing required to make the time series stationary, which is also known as the degree of differencing.

Mathematical formula for the AR and MA models:

Auto Regressive (AR only) model is one where Yt depends only on its lags. That is, Yt is a function of the 'lags of Yt'.

$$Yt = \alpha+\beta1Yt-1+\beta2Yt-2+..+\beta pYt-p+\varepsilon1 \_ \_ \_ \_ \_ \_ \_ \_ \_ \quad (1)$$

where, Yt–1 is the lag 1 of the series, β1 is the coefficient of lag 1 that the model estimates, and α is the intercept term, also estimated by the model.

$$Yt = \alpha+\varepsilon t+\varphi1\varepsilon t-1+\varphi2\varepsilon t-2+..+\varphi q\varepsilon t-q\_ \_ \_ \_ \_ \_ \_ \_ \_ \quad (2)$$

where the error terms are the errors of the auto-regressive models of the respective lags. The errors ϵt and $\epsilon t-1$ are the errors from the following equations:

$$Yt = \beta1Yt-1+\beta2Yt-2\cdot\cdot+\beta0Y0+\varepsilon tYt-1=\beta1Yt-2+\beta2Yt-3..+\beta0Y0+\varepsilon t-1\_ \_ \_ \_ \_ \_ \_ \_ \_ \quad (3)$$

Predicted Yt = Constant + Linear combination Lags of Y (up to p lags) + Linear Combination of Lagged forecast errors (up to q lags).

The frameworks need to be harmonized concerning the appropriate quandary since the models don't run as intended. An intensive perception of how the fundamental statistic models operate is required to harmonize these systems. The primary input variables to the ARIMA model are the moving average elements, the auto-regressive elements including the utmost degrees of difference. A standard interpreter does not apprehend how to monitor these systems to evade this operation and this sort of system is too laborious to accumulate and estimate. [27].

The Prophet Forecasting Model:

A perishable model comprising of three principal elements, namely trend, seasonality, and holidays is practised in the prophet forecasting model [14]. They're consolidated within the subsequent equation:

$$y(t)=g(t)+s(t)+h(t)+\varepsilon t(6) \_ \_ \_ \_ \_ \_ \_ \_ \_ (4)$$

g(t): section-wise linear or logistic growth curve used to model non-periodic fluctuations in the statistics.

s(t): periodical variations, which can be weekly or yearly seasonality.

h(t): impacts of holidays provided by the user with variable schedules.

єt : error term approximations for any significant changes not implemented by the model [18].

FB prophet attempts to readjust numerous linear and nonlinear functions of time by utilizing it as an independent variable. Exponential smoothing, as well as a prophet, practice the same strategy of modelling seasonality as a supplement component. In exponential smoothing, the prevailing observations are given more weight in forecasting compared to the earlier observations, since exponentially decreasing weights are ascribed due to the emergence of the observations [26].

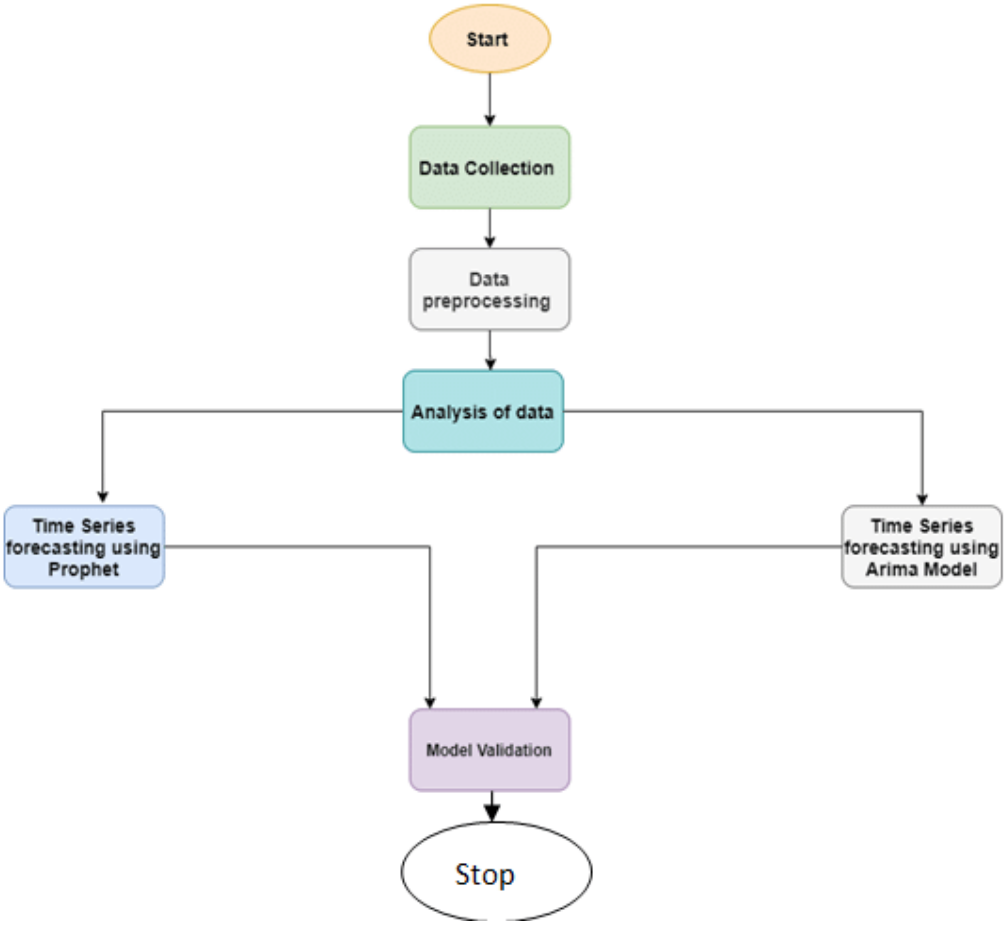

Figure 3.4: Flowchart of ARIMA model and Facebook Prophet

And for the execution of python code, we have used Google Colab.

Evaluation Parameters:

1) Root Mean Square Error (RMSE):

RMSE is the variance or the root of the residuals, where residuals tell how distant the regression curve is from the precise data points [21]. It's the measure of how the residuals disperse around the line of best fit. It will easily be deciphered because its units match the output units. Again, this is usually negatively-oriented and an inferior RMSE value improves the model performance [22].

$$RMSE = \sqrt{\{1/n \sum (y_i - y_i`)\}} \_ \_ \_ \_ \_ \_ \_ \_ \_ (5)$$

1) R-Squared Score (R^2):

R-squared (R^2) score is a metric of confidence that is easy to compute and intuitive to interpret [23]. It is the degree of how imminent a data point fits the linear regression; hence it tells us how good the regression line predicts the real values. It is the coefficient of determination that provides the measure of variation that has been demonstrated by the self-standing variables in the model. Furthermore, it gives us goodness-of-fit of the model, plus has a score that always prevails between 0 and 1 (0 and 100%). Due to inappropriate fit accompanied by the choice of an erroneous model, the values tend to lie outside the scope of the data. If R2 = 0.92, for instance, a 92% increase in the expense of fuel is due to the increase in the distance traversed. An R2 score of 1 symbolizes that the regression predictions fit the data faultlessly about its mean. A score of 0 indicates that none of the predictions fit the data about its mean. Higher the R2 score, better the model performance. R2 monotonously improves with the increase in the number of variables, but never diminishes. The formula to find the R-squared (R2) is given by [24]:

R^2= Variance Explained by Model /Total Variance_ _ _ _ _ _ _ _ (6)

CHAPTER 4

Experimental Results and Discussion

4.1 Data Visualization

Visualization is the graphic representation of data through the use of pictorial design. The goal is to make a visual easy to comprehend and presentable. [5] The main purpose of Coronavirus data-visualization is to observe information clearly and effectively using different graphical presentations. It is a useful medium for examining, understanding, and transmitting information because it has several possible uses in the domain.

4.1.1 Data Visualization libraries in python

1. **Matplotlib:** Matplotlib is a Python library used for plotting the beautiful and attractive Graphs. In data science visualization is the important step. By using visualization, we can easily understand than how data is split. It is a cross-platform library for making 2D plots from data in arrays. It provides an object-oriented API that helps in embedding plots in applications using Python GUI toolkits. Likewise, it can be used in Python and IPython shells, Jupyter notebook and web application servers also. [11]
2. **Seaborn:** Seaborn is a Python visualization library built on top of matplotlib to plot graphs. For drawing attractive statistical graphics, it provides a high-level interface. Seaborn is more integrated for working with Pandas DataFrames. [4]
3. **Plotly:** Plotly is built on d3.js and stack.gl. It is a high-level, declarative charting library. It ships with 20 chart types, including 3D charts, statistical graphs, and SVG maps. It supports a wide range of statistical, scientific, financial, geographic, and 3-dimensional visualizations.
4. **Folium:** Folium is a powerful Python library that helps with mapping geospatial data to create several types of Leaflet maps from scratch with Python. builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the Leaflet.js library and creates a map in a separate HTML file. [12]

4.1.2 Data Visualization Result

In this research, we have imported many data visualization libraries such as Matplotlib.pyplot, Seaborn, Plotly.express, Folium, Plotly.graph_objects. Through those libraries, we were able to analyze and produce some meaningful graphical data representation from COVID-19 datasets. Researchers identified the following finding based COVID-19 datasets:

The following figure 4.1 shows COVID-19 situation of South Asian countries over time according to confirmed cases, death cases and recovered cases.

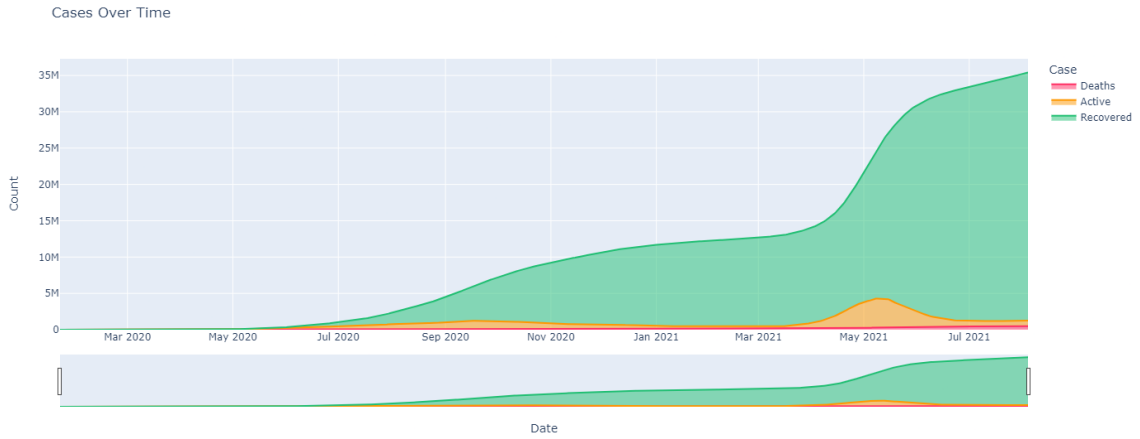

Figure 4.1: COVID-19 cases over time in South Asian countries

Here it is observed that the line gradually increased, but it was going straight comparatively during the winter (up to early March) and again rises high when summer (from late March) came.

The following figure 4.2 shows choropleth map of COVID-19 situation of South Asian countries over time according to confirmed cases.

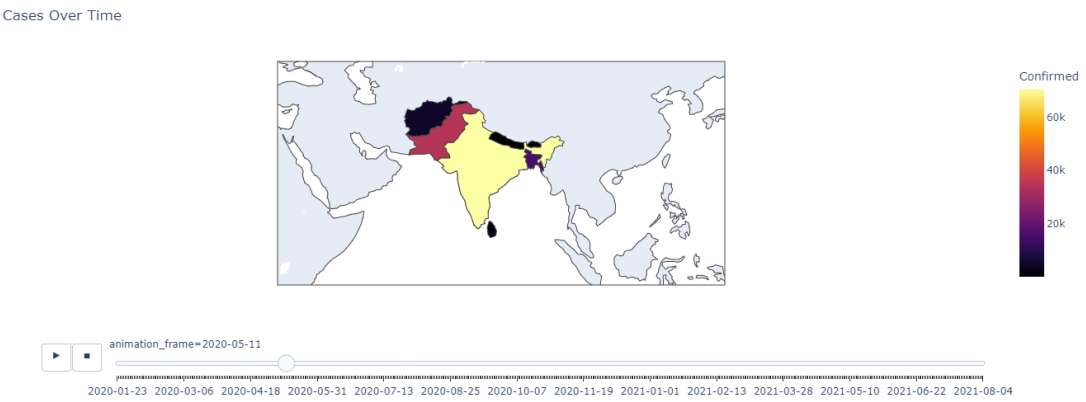

Figure 4.2: Choropleth map visualizing the COVID-19 cases over time in South Asian countries

In this choropleth map, the color of the countries will be changed overtime as shown in the color bar in the right portion of the figure, according to the confirmed cases. In the end it can be observed that India got the highest number of confirmed cases and then Bangladesh, Pakistan, Nepal, Sri Lanka, Afghanistan, Maldives and Bhutan. Here the confirmed cases were more depending on the number of populations.

The following figure 4.3 shows COVID-19 situation of South Asian countries into four phases (divided by 4 months approximately) according to confirmed cases, death cases and recovered cases visualized using scatter plots.

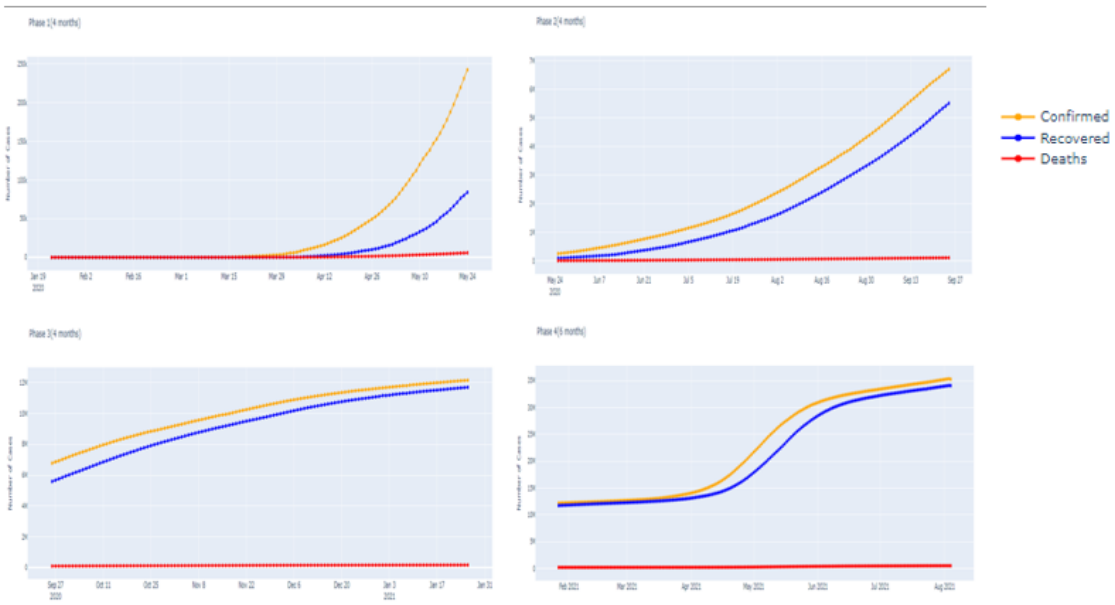

Figure 4.3: Four phases of COVID-19 in South Asian countries

The following figure 4.4 shows COVID-19 situation of South Asian countries into four phases (divided by 4 months approximately) according to confirmed cases and visualized using bar graphs.

Figure 4.4: Four phases of Confirmed cases of COVID-19 in South Asian countries

From the above visualization, we can see that the bar of India is always high among the rest of the countries. Situation of Afghanistan, Bangladesh and Pakistan was high in the first phase, but later it seems to be low compared to India's situation and other countries' situation is almost negligible compared to mentioned countries.

The following figure 4.5 shows COVID-19 situation of South Asian countries into four phases (divided by 4 months approximately) according to recovered cases and visualized using bar graphs.

Figure 4.5: Four phases of Recovered cases of COVID-19 in South Asian countries

The following figure 4.6 shows COVID-19 situation of South Asian countries into four phases (divided by 4 months approximately) according to confirmed cases and visualized using bar graphs.

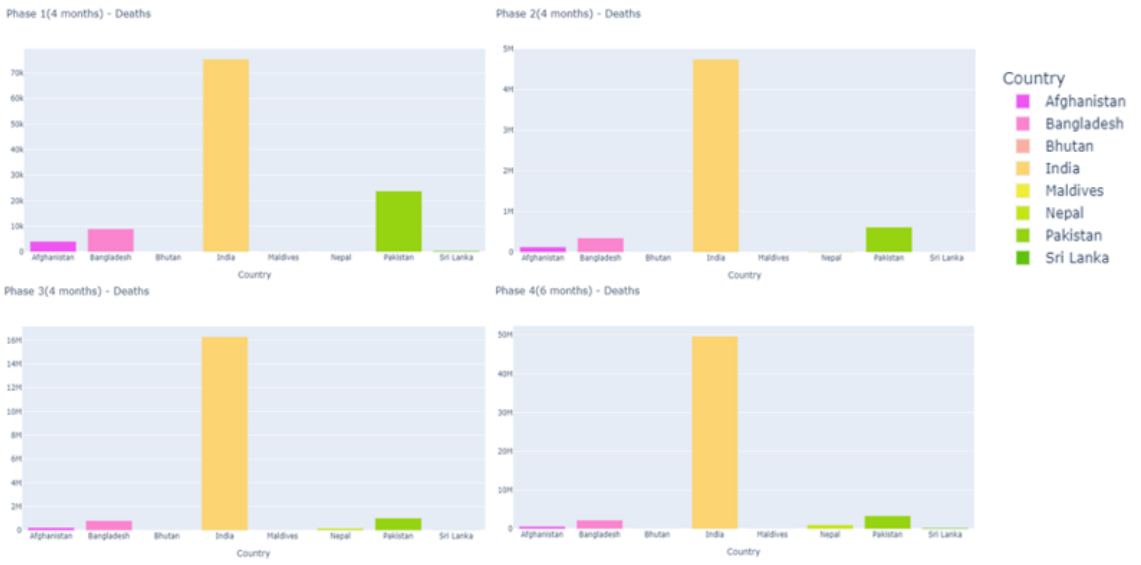

Figure 4.6: Four phases of Death cases of COVID-19 in South Asian countries

According to the Ministry of Public Health, on August 5, 2020, around a third of the population was affected by COVID-19, that's roughly 10 million. Cases have gone significantly down from there, and we can see a recovery in the economy as well. Kabul Province had the highest number of COVID-19 cases for Afghanistan. From our

visualization we also saw that for per 100 cases they were leading the chart for death cases among the South Asian countries. [11]

Bangladesh being a densely populated country did suffer a lot from COVID-19 and there were significant number of loss of lives in this country. The government was quick enough to declare multiple lockdowns to reduce the spread of the virus. But still they peaked at one time to be the highest new cases in Asia. The outbreak was severe and even crossed the number of cases in China. Death were reported daily. Bangladesh needed to ready itself, by establishing a demand forecasting model for Medical-grade Oxygen at the earliest when the spread was widespread for the virus with the coordinated efforts of Department of Public Health Engineering (DPHE); and Institute of Epidemiology, Disease Control and Research (IEDCR). With these combined efforts, the rates for confirmed cases have gone down, with some late death cases being confirmed on recent variants of COVID-19. [8][9]

Bhutan did a great work for reducing the spread of COVID-19 in their country compared to other South Asian countries. Their government was fast to implement appropriate measures to stop the spread of the virus. On recent variants, their citizens were affected, but the mortality rate remained low. And that is visible in our visualization. [19]

India is another densely populated country from South Asia and for that, their people had to suffer from lack of medical facilities as the country doesn't have adequate infrastructure to support every affected citizen. Government was quick to take appropriate measures to tackle the situation but when the spread was uncontrollable people were getting affected daily and death rates kept increasing in India. With vaccination done on a mass level, the situation was brought back to normal and from there we are seeing a declined number of cases in terms of deaths and confirmed cases. India was leading the daily confirmed cases and active cases for a while in comparison to global stats. Oxygen shortage was a big issue for them. Vaccines were given to elderly people to reduce the deaths. And from there things started to slow down. [8][9]

Although the death rates were comparatively lower for the Maldives, at one point they were the fastest growing country in terms of infection for COVID-19. Things have slowed down from there after the government took appropriate measures to handle the situation.

Their tourism sector was hampered heavily due to the lockdowns, which caused massive economic losses. Mass vaccination played a vital role to minimize the death cases. [13]

COVID-19 caused serious economic damage to Nepal as they are dependent on tourism heavily and for the lockdowns that were necessary to control the spread of the virus, they had to take losses on the economy. Bagmati Province and Kathmandu were hit the most. [14]

Pakistan is another densely populated country of South Asia where the effects of COVID-19 were clearly visible. Karachi, Lahore, Islamabad and Peshawar were hit the hardest by the virus with most cases. After the government took strict measures, the number started to decline along with the vaccination programs. [8][9]

Sri Lanka is another country that is dependent on tourism for their national economy, so the pandemic caused the downfall of their economy when it started to hit the country. The government took measures to control the situation with measures like lockdown, but they couldn't do a proper rebound like other countries of South Asia after the adverse effects of COVID-19 started to deteriorate with mass vaccination. They are currently facing national bankruptcy. The government in power is already trying to flee the country as global leaders try to help stabilize the national economy. COVID-19, maybe one of the root causes of downfall of Sri Lanka. [15]

The following figure 4.7 shows COVID-19 situation of South Asian countries according to confirmed cases, death cases and recovered cases separately and visualized using line plots.

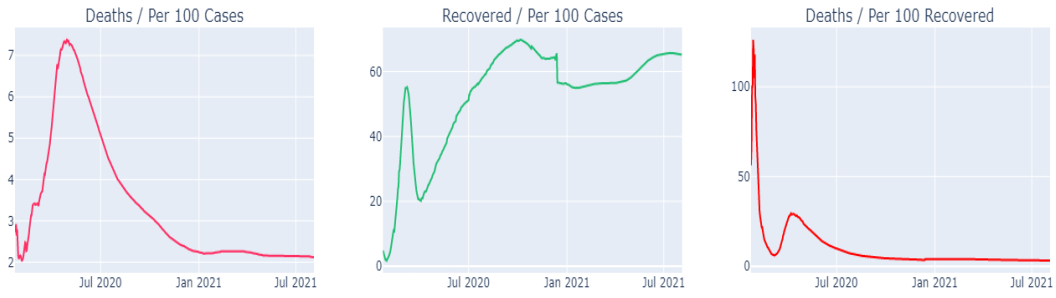

Figure 4.7: Visualization on per 100 cases in all South Asian countries

From the above plots, we can see that the line of Deaths/Per 100 cases was high during the early stage of COVID-19, but in the later stage the line went downward. Again, in Recovered/Per 100 cases the line was in the downside but in the later stage it went high. Thus, it can be seen that COVID-19 was much deadlier in the earlier stage than the later stage. Since there were lots of variants of COVID, thus we can say that the variants in the later stages aren't deadlier as the early variants.

The following figure 4.8 shows COVID-19 situation of each country of South Asia according to recovered/per 100 cases, deaths/per 100 cases, deaths/per 100 recovered cases and cases/million population separately and visualized using bar diagrams.

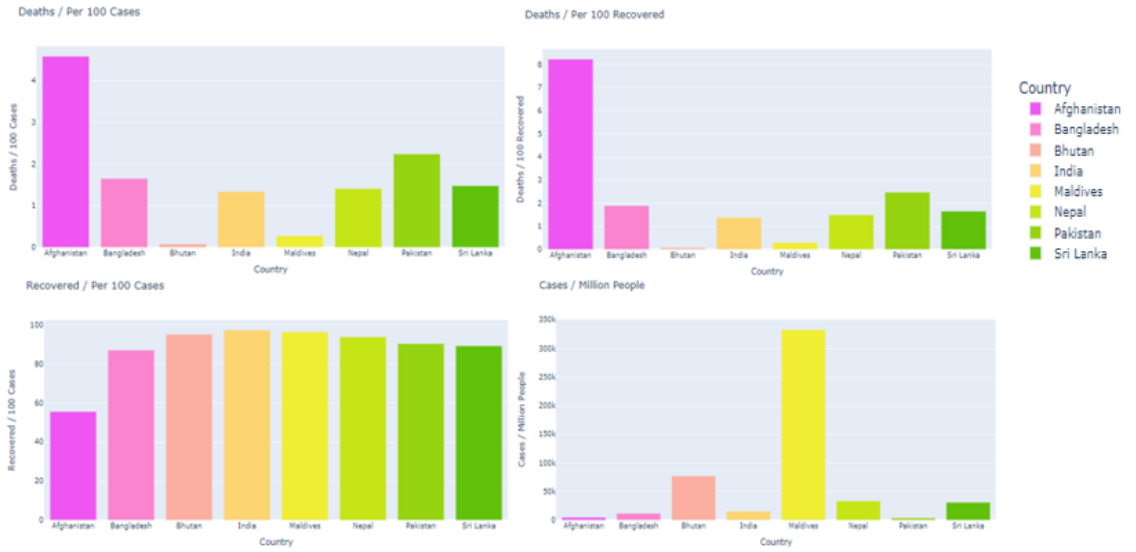

Figure 4.8: Visualization on per 100 cases country wise

From the above bar plots, we can see that the bar in Recovered/Per 100 cases of Afghanistan is lower than the other countries again the bar in Deaths/Per 100 cases is the highest than the other countries, and thus we can say that among all other South Asian countries, Afghanistan couldn't tackle the situation like other countries. [16]

4.2 Prediction Model

The ARIMA & Facebook Prophet models have been used on our collected dataset for predicting the result of confirmed, recovered and death cases for 30 days. The data which

we have used is provided by Johns Hopkins which gives the cumulative case count for each country of South Asia on each day from 22-01-2020 to 08-05-2021, total 473 days.

Before training an ARIMA model on the data, the data needs to be checked for stationarity, meaning that the data has stable statistics, such as constant means and variances, over time. This is because, otherwise, the timeseries is essentially equivalent to a white noise, carrying little information. Stationarity helps to ensure that behavior observed in a particular window of the data is likely to happen again, which is why many statistical models depend on it. To assist in finding p and q, ACF and PACF plots were created. The autocorrelation function gives the effect that all previous points in time together have on the current point. Hence, the moving average term q can be derived. [20][21]

The partial autocorrelation function gives the *direct* effect that each of the previous points have on the current observation. After that, a stepwise search is run to iterate over p, d and q values and select the best parameter pair based off of the AIC score, which estimates out-of-sample prediction error. A low AIC score is generally considered better and indicates how well a model fits the data without overfitting. Datasets with seasonality in them can be forecasted accurately using prophet. It's an open source software that was released by Facebook's core Data Science Team. It can deal pretty well with yearly, weekly, & daily seasonality along with holiday effects by removing these trends. [22][23]

TABLE 1: ARIMA & Facebook Prophet models for South Asian countries with evaluation parameters

Country	**Model**	**Case Type**	**RMSE**	**R^2**
Afghanistan	ARIMA (4,2,2)	Confirmed	859.53	0.722
	Facebook Prophet		364.73	0.92
	ARIMA (3,1,2)	Recovered	615.93	0.684
	Facebook Prophet		301.87	0.93
	ARIMA (2,2,4)	Deaths	19.27	0.852
	Facebook Prophet		13.86	0.91
Bangladesh	ARIMA (5,2,4)	Confirmed	100113.42	0.651

	Facebook Prophet		15638.63	0.90
	ARIMA (0,2,1)	Recovered	40305.70	0.723
	Facebook Prophet		13420.05	0.92
	ARIMA (1,2,2)	Deaths	362.99	0.751
	Facebook Prophet		238.01	0.91
Bhutan	ARIMA (2,2,1)	Confirmed	112.86	0.771
	Facebook Prophet		32.56	0.88
	ARIMA (1,2,2)	Recovered	66.33	0.725
	Facebook Prophet		17.58	0.89
	ARIMA (0,1,0)	Deaths	0	1
	Facebook Prophet		0.17	0.848
India	ARIMA (5,2,5)	Confirmed	1732463.99	0.712
	Facebook Prophet		824821.07	0.90
	ARIMA (2,2,2)	Recovered	1736055.72	0.722
	Facebook Prophet		523802.26	0.93
	ARIMA (2,2,2)	Deaths	22244.70	0.752
	Facebook Prophet		638.66	0.91
Maldives	ARIMA (1,2,1)	Confirmed	1938.62	0.753
	Facebook Prophet		423.44	0.94
	ARIMA (1,2,2)	Recovered	186.06	0.90

	Facebook Prophet		72.41	0.95
	ARIMA (0,2,2)	Deaths	3.65	0.855
	Facebook Prophet		1.11	0.90
Nepal	ARIMA (5,2,1)	Confirmed	39536.44	0.822
	Facebook Prophet		8435.54	0.88
	ARIMA (2,2,1)	Recovered	7741.32	0.812
	Facebook Prophet		1757.18	0.89
	ARIMA (0,2,1)	Deaths	88.90	0.721
	Facebook Prophet		99.88	0.87
Pakistan	ARIMA (4,2,0)	Confirmed	6647.14	0.981
	Facebook Prophet		12717.199	0.97
	ARIMA (1,2,2)	Recovered	6647.14	0.981
	Facebook Prophet		7807.55	0.94
	ARIMA (2,2,1)	Deaths	525.27	0.783
	Facebook Prophet		241.37	0.94
Sri Lanka	ARIMA (3,2,4)	Confirmed	8886.25	0.741
	Facebook Prophet		1709.09	0.93
	ARIMA (0,2,1)	Recovered	2133.78	0.764
	Facebook Prophet		1599.28	0.90
	ARIMA (3,2,5)	Deaths	37.11	0.842

	Facebook Prophet		10.16	0.89
Sample Size: 473 days Time Frame: 22-01-2020 to 08-05-2021				

Confirmed Cases Forecasting (Facebook Prophet):

The following figure 4.9 shows a slow inclination for the coming months, meaning confirmed cases were going down. However, the cases reached three folds higher than it was in the beginning. The country has faced a slight peak at the end, but it came to lower position compared to most countries.

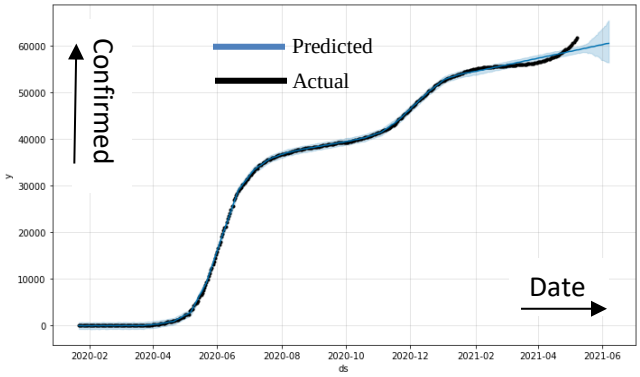

Figure 4.9: Confirmed cases forecasting of Afghanistan

The following figure 4.10 shows a similar trend, with confirmed cases going down. However, the path followed a very uniform increase in the number of cases. Similar pattern is observed in death & recovered case. Although the country is pretty densely populated, the government did a good job of handling the situation well with early lockdowns.

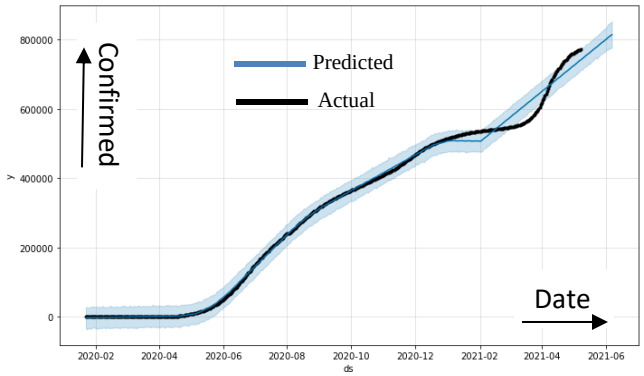

Figure 4.10: Confirmed cases forecasting of Bangladesh

The following figure 4.11 shows very low increase for the confirmed cases here. The increase in the cases was gradual, but a sudden increase has been observed twice within the given time span. Confirmed cases stayed pretty low compared to the situation of COVID-19 in the world. We see a similar pattern when we look at the death & recovered graphs for the timeframe, we are running our studies on.

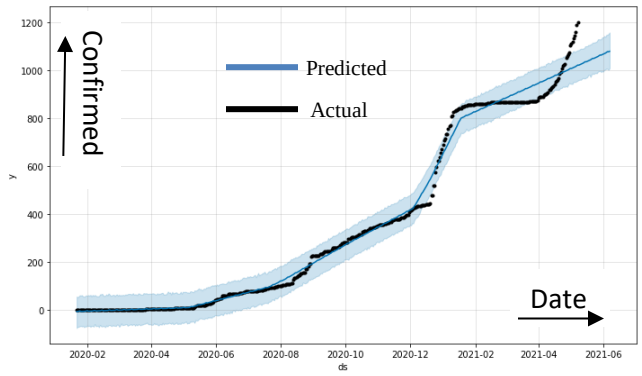

Figure 4.11: Confirmed cases forecasting of Bhutan

The following figure 4.12 shows a comparatively higher number of confirmed cases for the coming months. The country has faced a fast-paced increase in the number of cases, which grew over the time. One of the major reasons for this is the population being high & densely populated in many areas of the country.

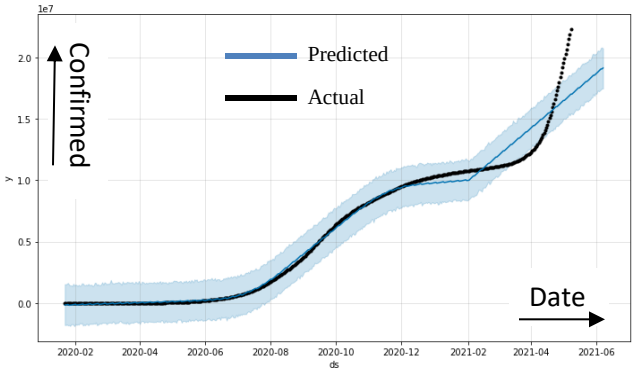

Figure 4.12: Confirmed cases forecasting of India

The following figure 4.13 shows a low increased number of confirmed cases. There is no sign of sudden increase of cases and the cases follow a very slow increase in number compared with the given time. It followed this steady trend line for a while before coming down significantly.

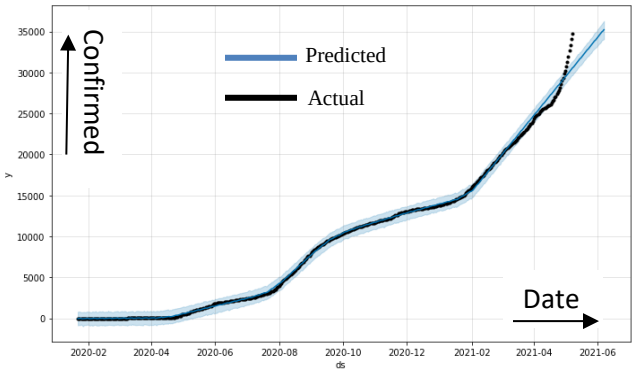

Figure 4.13: Confirmed cases forecasting of Maldives

The following figure 4.14 shows significant jump in confirmed cases for the following months. The rise of confirmed cases was very low as shown in the figure which took a steep turn upwards and maintained an increased number of cases which was stable for a while but faced with another sudden rise within a short span of time.

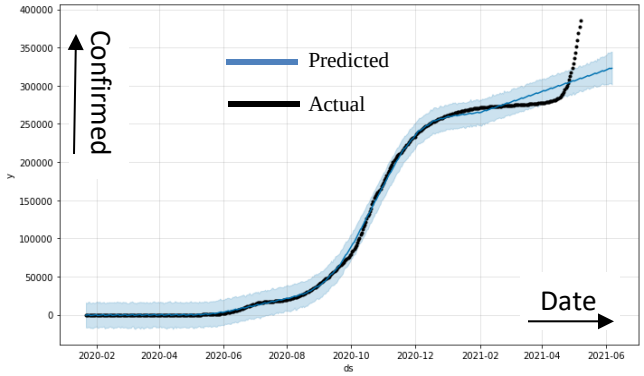

Figure 4.14: Confirmed cases forecasting of Nepal

The following figure 4.15 shows a comparatively higher number in terms of confirmed cases. The increase in number of cases in very clear as the fast-growing cases goes upwards.

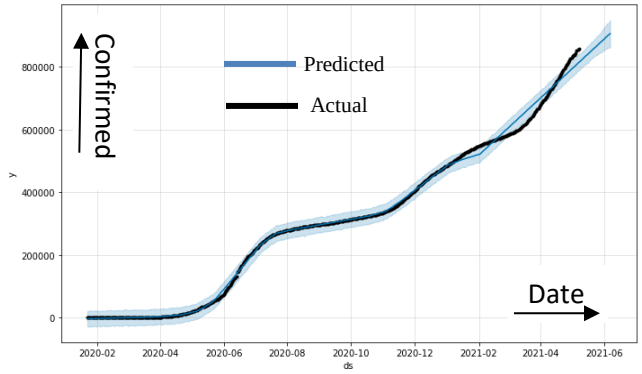

Figure 4.15: Confirmed cases forecasting of Pakistan

The following figure 4.16 shows a slow increase in confirmed cases. Though the figure shows very slow confirmed cases from 2020-02 to 2020-10, a sharp increase in the number of cases has been observed, which demonstrates a high increase within a short span of time.

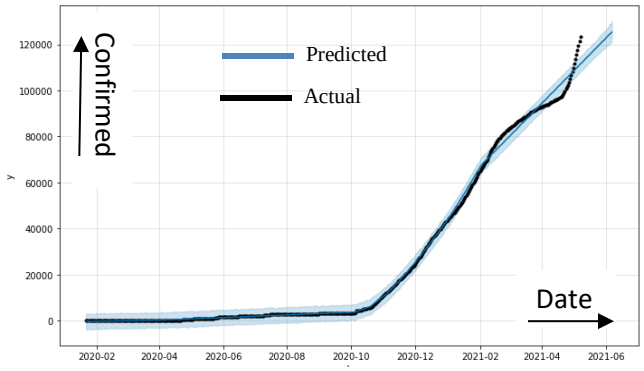

Figure 4.16: Confirmed cases forecasting of Sri Lanka

Recovered Cases Forecasting:

The following figure 4.17 shows a low number of recovered cases. The recovered cases are low compared to the time span and the number of cases. This is mainly due to the population being not so high. The initial lockdowns might have helped.

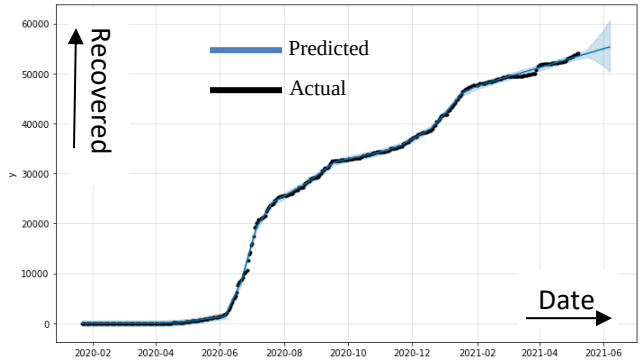

Figure 4.17: Recovered cases forecasting of Afghanistan

The following figure 4.18 shows also a low number of recovered cases. Due to the gradual increase number of confirmed cases, conversely the steady number of recoveries is observed here.

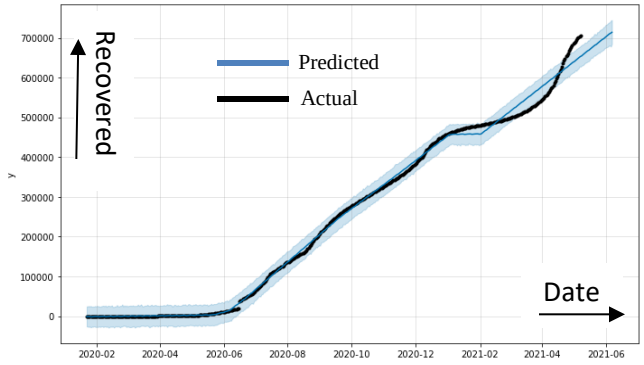

Figure 4.18: Recovered cases forecasting of Bangladesh

The following figure 4.19 shows a low number of recovered cases due to less confirmed cases previously seen. Since the number of confirmed was low, it maintained an equilibrium given a low number of recovered cases. With a low number of confirmed cases, the country had a steady increase in the number of recovered cases, but it's definitely low when compared to most countries. Also, the population is not that large for them.

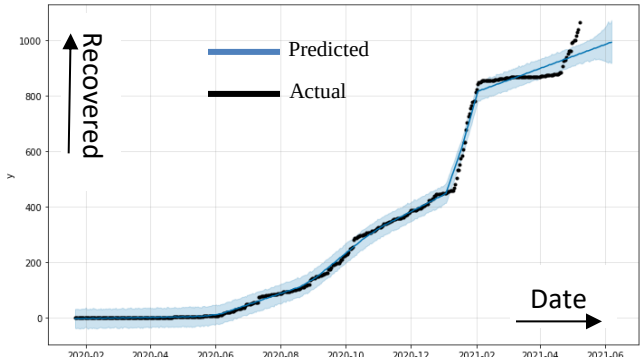

Figure 4.19: Recovered cases forecasting of Bhutan

The following figure 4.2.20 shows a significantly higher number of recovered cases. Compared to the number of confirmed cases, the recovered cases are higher. Due to proper medical facilities, the number of recovered cases increased over the time for the country. Foreign aids definitely played a role in here.

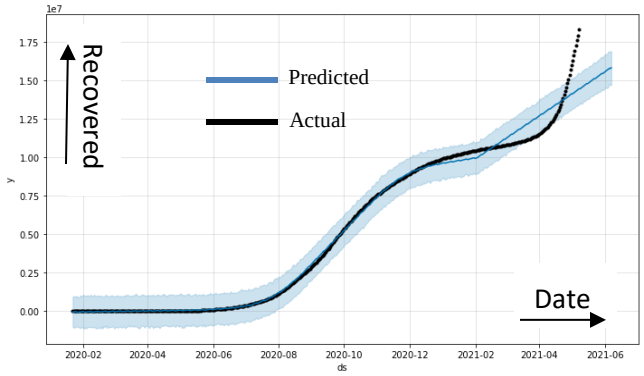

Figure 4.20: Recovered cases forecasting of India

The following figure 4.21 shows a low number of recovered cases as well. The low number of cases recovered has been observed here within 2021-06 confirmed cases rose so high that the recovered cases were very low.

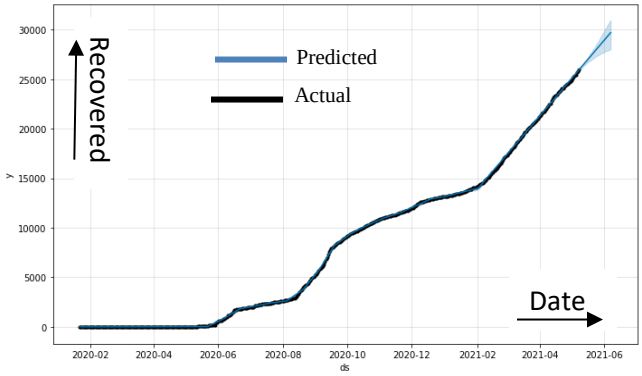

Figure 4.21: Recovered cases forecasting of Maldives

The following figure 4.22 shows low number of recovered cases compared to confirmed cases. The gradual increase of confirmed cases reciprocated in a lower number of recovered cases, which was seen in the case of Nepal.

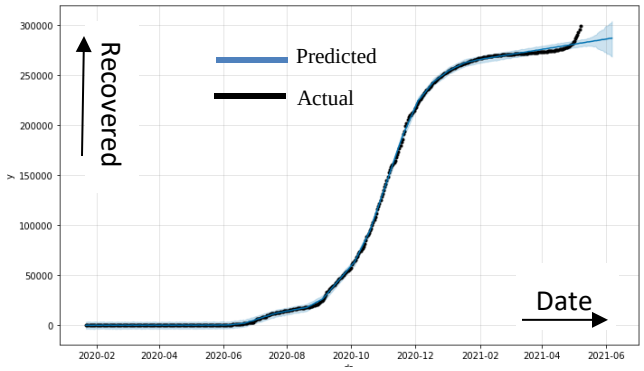

Figure 4.22: Recovered cases forecasting of Nepal

The following figure 4.23 shows a higher number of recovered cases compared to confirmed cases. Despite the increase of confirmed rates by three-fold, the recovered cases aren't low either.

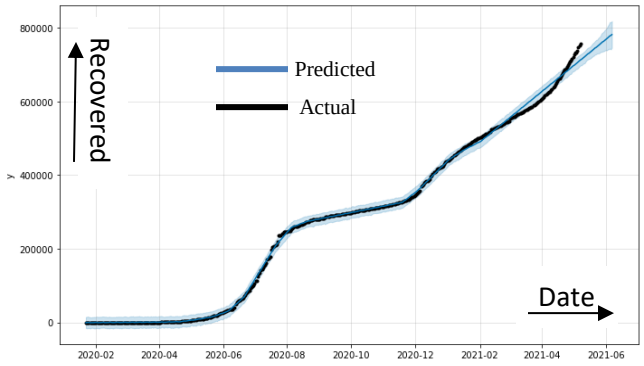

Figure 4.23: Recovered cases forecasting of Pakistan

The following figure 4.24 shows a significantly higher number of recovered cases compared to confirmed cases. Due to the significantly low number of confirmed cases till 2020-10, it significantly resulted in a higher number of recovered cases.

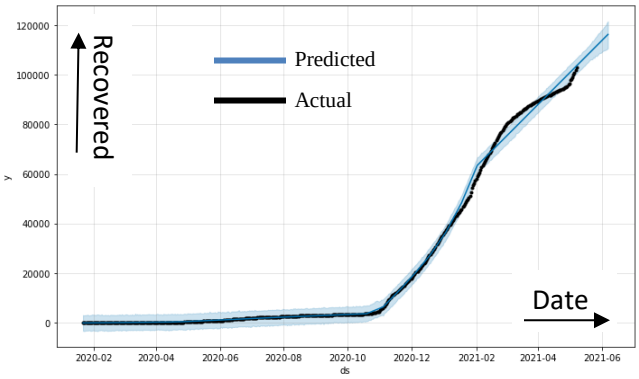

Figure 4.24: Recovered cases forecasting of Sri Lanka

Death Cases Forecasting:

The following figure 4.25 shows a significantly low number of death cases. The death rate stayed within a range that's not very high over the period of time. Again, we can say this is due to the fact that the population is not that large for the country & lockdowns helped immensely to keep the deaths low during the peak periods.

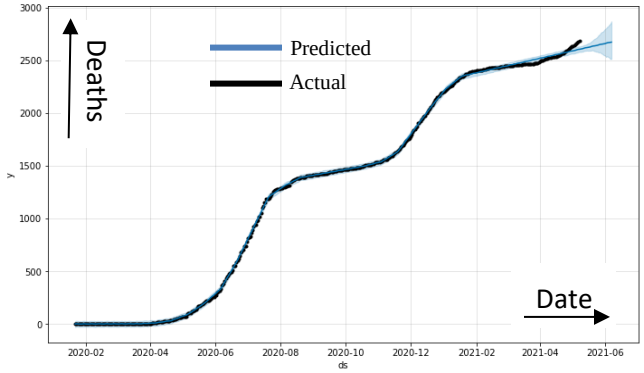

Figure 4.25: Death cases forecasting of Afghanistan

The following figure 4.26 shows similar low number of death cases. We see a slight increase of death in the end of the time frame we are running our charts on. After that, the deaths have gone lower as the country becomes more aware of the virus & overall situation.

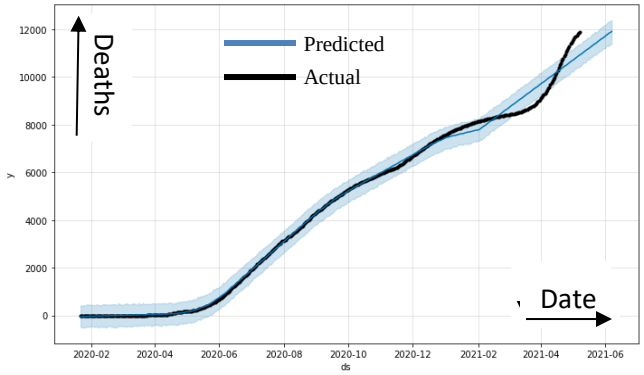

Figure 4.26: Death cases forecasting of Bangladesh

The following figure 4.27 shows an extremely low number of death cases. The country has the lowest number of deaths. This is due to strict decisions taken by the government to enforce lockdown when the virus first appeared in the country. This helped to keep the death rate extremely low.

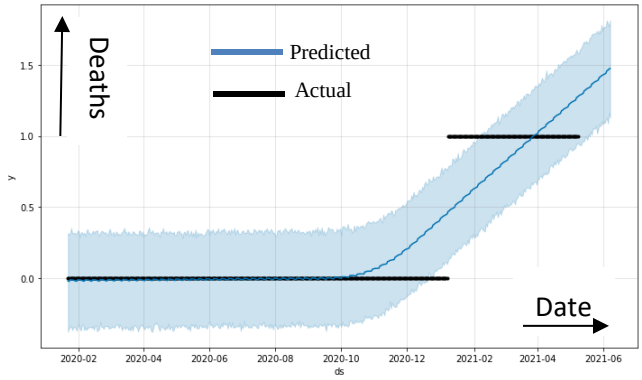

Figure 4.27: Death cases forecasting of Bhutan

The following figure 4.28 shows comparatively higher number of death cases. Being a densely populated, the number of deaths stayed pretty high for them throughout the waves of the virus. Although we see a major peak for the at the end of the time frame when we look at the population to death ratio, it's actually within a good threshold for them. With proper medical facilities, the number came down eventually.

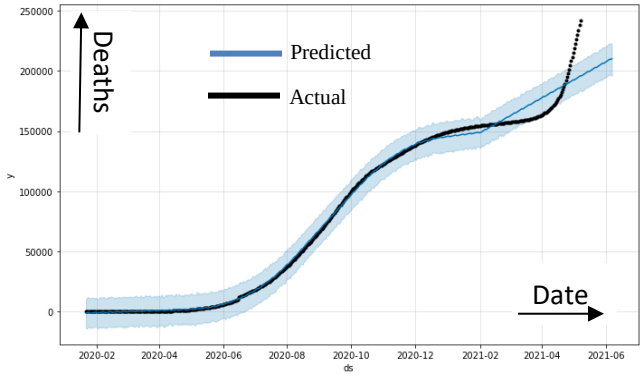

Figure 4.28: Death cases forecasting of India

The following figure 4.29 shows a low number of death cases. We see a similar case like Bhutan for the country where the deaths stayed pretty low throughout the time frame. This is again due to not so high population & enough medical facilities available for the people in the country. Initial lockdowns also help immensely for keeping the number of deaths low.

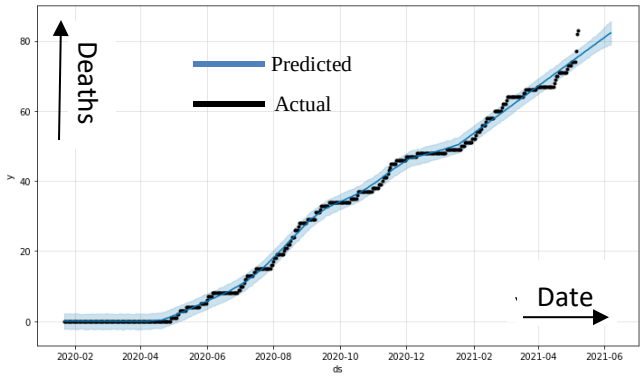

Figure 4.29: Death cases forecasting of Maldives

The following figure 4.30 shows a low number of death cases as well. With low population, the country faced significantly low number of deaths. Public awareness & initial lockdowns definitely played a major role to keep this death cases low throughout the time frame we are studying on.

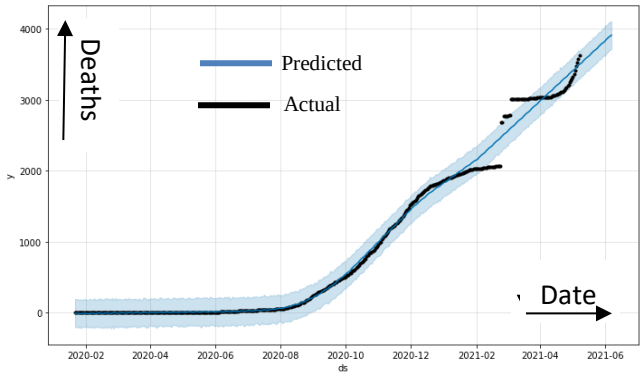

Figure 4.30: Death cases forecasting of Nepal

The following figure 4.31 shows a slightly increased number of death cases compared to other countries here. With dense population, we can see a comparatively high number of deaths for them. At the end of the time frame the number has definitely dropped with more medical facilities being available to people & overall public awareness.

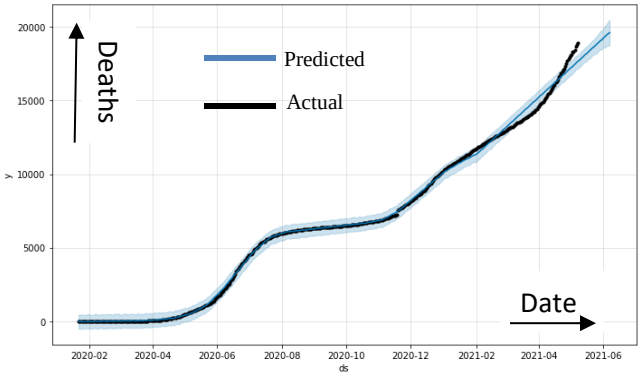

Figure 4.31: Death cases forecasting of Pakistan

The following figure 4.32 shows significantly higher number of death cases in the following months with a flat curve initially. We see a low number in the initial phase, but at the later phase we see pretty good peaks for the deaths. It came out eventually with proper steps taken by the government along with medical support & public awareness.

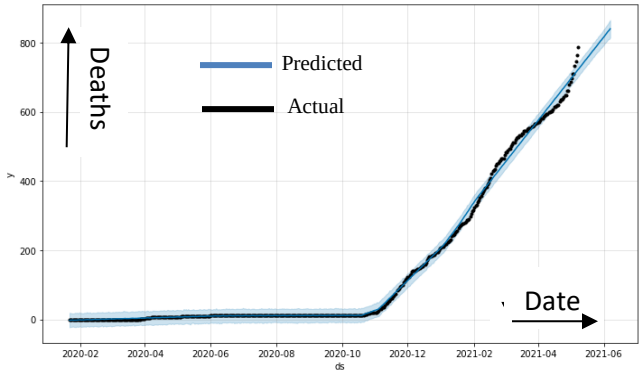

Figure 4.32: Death cases forecasting of Sri Lanka

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Understanding COVID-19 is really important for everyone, not only South Asia. But being developing nations with mostly dense population, pandemics like COVID-19 can do substantial damage to the economy & lives of people. Starting from lack of proper medical facilities to having not enough hospital beds for affected people, COVID-19 really taught the whole world why we need to put more time & energy in research, especially in our medical sector. As viruses mutate themselves, we need to better prepare for the future that holds before us. This paper is centered on that. We focused on the COVID-19 situation in South Asia, how the virus shaped out throughout different phases. While most countries in South Asia got crushed by COVID-19 & took a massive hit on the economy, most of them didn't have enough medical facilities in the first place to support the affected people, especially the elderly ones who needed ICUs. Even in Bhutan where they had minimal loss of lives at first, they got hit at the end phase & a similar situation was seen there. We need more papers on COVID-19 so that the governments become more aware & people start demanding their security especially in the medical sector.

Statistical data & understanding the trends of the virus can give us the edge that we need to prepare ourselves for the next wave of the virus or any similar virus. Doing more research on such topics can really help us minimize the risk of losing more lives of people in the future. Data science have progressed so much in past few decades & it is really important to use it to our advantage right now.

5.2 Impact on Environment

While COVID-19 is a virus that has been affecting mostly humans. The impact on environment can also be seen as well. Due to lockdowns millions of people were bound to stay at home & work from home. These reduced the carbon emissions from vehicles as fewer vehicles were in the streets. Thus, improving the air quality of the environment. Water pollution also decreased in many places as mills & factories were shut down due to

the pandemic. So, we can definitely say that the environment got the better side of the COVID-19 & healed itself.

5.3 Ethical Aspects

COVID-19 brought up serious health awareness discussion on the table. Governments from around the globe were bound to go back to the drawing board to reevaluate their medical sectors & health facilities that they hold for their citizen. Population health maximization, justice for everyone, no harm principle, trustworthiness & solidarity were most discussed topics for them. New hospitals were also set up in many countries due to the pandemic. So definitely some significant discussions & changes were made.

5.4 Sustainability Plan

Post COVID-19 situation made the governments around the globe take some policy changes in their medical & health sectors. South Asia already falls under the SDG policy, so, for them COVID-19 was a good reminder why those goals are necessary for them to achieve. If we take a look into other countries, they also made significant changes to their policies that include health & medical sector because everyone saw the adversity of COVID-19 & how such pandemic can disrupt the national economy immensely. Work from home also got popular for many professionals after the pandemic & companies are encouraging their employees to take that option if they opt for it. More or less everyone implemented some changes to tackle with any future pandemic.

CHAPTER 6

Conclusions and Future work

6.1 Conclusions

In this research paper, we have presented a data visualization of COVID-19 dataset in a multidimensional and multi-faceted way. The research paper aimed at providing insights, new directions and opportunities for research in the field of COVID-19 Data visualization and analytics especially for the South Asian countries**.** We have also predicted the spread of COVID-19 for 30 days of each country of South Asia to understand the curve/pattern they might follow in the upcoming days outside our sample data.

Our comparative study definitely showed us that countries which were densely populated didn't have adequate health and medical facilities to face this pandemic. Things could have been tackled better if infrastructure was built over time and not rushed during pandemic. Countries with smaller populations did face less severity of the virus this time, but they shouldn't neglect their infrastructures as well. COVID-19 was an awakening message for everyone, not only South Asian countries. As developing countries, we need to be prepared for the next pandemic in terms of infrastructural developments as our economy is dependent on the human resource or people of our countries. The governments and policymakers should be more aware of this and work accordingly. Public awareness can be raised as well in order to tackle such situations in the future.

6.2 Future Work

In the future, we will look at how to improve the performance of prediction, and we will try to improve the prediction time too. We are also planning to update this study with more analyses and cases continuously by fine-tuning the prediction on other countries of the world. We will also try to study on detailed data as per different variants of COVID and add features on data like sex, age and symptoms.

REFERENCES

[1] Malki, Z., Atlam, ES., Ewis, A. *et al.* The COVID-19 pandemic: prediction study based on machine learning models. *Environ Sci Pollut Res* 28, 40496–40506 (2021).

[2] Sahai, A.K., Rath, N., Sood, V. and Singh, M.P., 2020. ARIMA modelling and forecasting of COVID-19 in top five affected countries. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, *14*(5), pp.1419-1427.

[3] Sun, J., Chen, X., Zhang, Z., Lai, S., Zhao, B., Liu, H., Wang, S., Huan, W., Zhao, R., Ng, M.T.A. and Zheng, Y., 2020. Forecasting the long-term trend of COVID-19 epidemic using a dynamic model. *Scientific Reports*, *10*(1), pp.1-10.

[4] Oberoi, A., and Chauhan, R. (2019). Visualizing data using Matplotlib and Seaborn libraries in Python for data science. International Journal of Scientific and Research Publications (IJSRP), 9(3), p8733. https://doi.org/10.29322/ijsrp.9.03.2019.p8733

[5] Shaito, M. and Elmasri, R., 2021, June. Map visualization using spatial and spatio-temporal data: Application to Covid-19 data. In *The 14th PErvasive Technologies Related to Assistive Environments Conference* (pp. 284-291).

[6] Ogundokun, R.O. and Awotunde, J.B., 2020. Machine learning prediction for Covid-19 pandemic in india. *medRxiv*.

[7] Alodat, I., 2021. Analysing and predicting COVID-19 AI tracking using artificial intelligence. *International Journal of Modeling, Simulation, and Scientific Computing*, *12*(03), p.2141005.

[8] Younus, U., 2021. The Impact of COVID-19 on South Asian Economies.

[9] Rasul, G., Nepal, A.K., Hussain, A., Maharjan, A., Joshi, S., Lama, A., Gurung, P., Ahmad, F., Mishra, A. and Sharma, E., 2021. Socio-economic implications of COVID-19 pandemic in South Asia: emerging risks and growing challenges. *Frontiers in sociology*, *6*, p.629693.

[10] Shetty, A. and Shetty, S., 2014. The impact of doctors per capita on the mortality rate in Asia. *Int. J. Med. Pharm. Sci*, *4*, pp.10-15.

[11] Shah, J., Karimzadeh, S., Al-Ahdal, T.M.A., Mousavi, S.H., Zahid, S.U. and Huy, N.T., 2020. COVID-19: the current situation in Afghanistan. *The Lancet Global Health*, *8*(6), pp.e771-e772.

[12] de Bengy Puyvallée, A. and Storeng, K.T., 2022. COVAX, vaccine donations and the politics of global vaccine inequity. *Globalization and Health*, *18*(1), pp.1-14.

[13] Rauf, M., 2021. Issues and Impact of COVID-19 on Socio-Legal Aspects of Human Life: A Case Study of Greater Male Region'in the Maldives. *Asian Journal of Research in Business and Management*, *3*(1), pp.70-79.

[14] Poudel, K. and Subedi, P., 2020. Impact of COVID-19 pandemic on socioeconomic and mental health aspects in Nepal. *International Journal of Social Psychiatry*, *66*(8), pp.748-755.

[15] Deyshappriya, N.P., 2020. Economic impacts of COVID-19 macro and microeconomics evidences from Sri Lanka. *Economic Impacts of COVID-19 Macro and Microeconomics Evidences from Sri Lanka (May 10, 2020)*.

[16] Essar, M.Y., Hasan, M.M., Islam, Z., Riaz, M.M.A., Aborode, A.T. and Ahmad, S., 2021. COVID-19 and multiple crises in Afghanistan: an urgent battle. *Conflict and Health*, *15*(1), pp.1-3.

[17] "Arima model - complete guide to time series forecasting in python -ml+"- https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python

[18] I. Yenidoğan, A. Çayir, O. Kozan, T. Dağ and Ç. Arslan, "Bitcoin forecasting using arima and prophet", 2018 3rd International Conference on Computer Science and Engineering (UBMK), pp. 621-624, 2018.

[19] S. J. Taylor and B. Letham, "Forecasting at scale", The American Statistician, vol. 72, no. 1, pp. 37-45, 2018.

[20] G. P. Zhang, "Time series forecasting using a hybrid arima and neural network model", Neurocomputing, vol. 50, pp. 159-175, 2003.

[21] W. Wang and Y. Lu, "Analysis of the mean absolute error (mae) and the root mean square error (rmse) in assessing rounding model", IOP Conference Series: Materials Science and Engineering, vol. 324, no. 1, pp. 012049, 2018.

[22] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance", Climate research, vol. 30, no. 1, pp. 79-82, 2005.

[23] J. Lupón, H. K. Gaggin, M. De Antonio, M. Domingo, A. Galán, E. Zamora, J. Vila, J. Peñafiel, A. Urrutia, E. Ferrer et al., "Biomarker- assist score for reverse remodeling prediction in heart failure: the st2-r2 score", International journal of cardiology, vol. 184, pp. 337-343, 2015.

[24] O. Renaud and M.-P. Victoria-Feser, "A robust coefficient of determination for regression", Journal of Statistical Planning and Inference, vol. 140, no. 7, pp. 1852-1862, 2010.

[25] Choudhary, A., 2018. Generate quick and accurate time series forecasts using Facebook's Prophet (with Python & R codes). *Analytics Vidya*.

[26] Poleneni, V., Rao, J.K. and Hidayathulla, S.A., 2021, January. COVID-19 prediction using ARIMA model. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 860-865). IEEE.

[27] Battineni, G., Chintalapudi, N. and Amenta, F., 2020. Forecasting of COVID-19 epidemic size in four high hitting nations (USA, Brazil, India and Russia) by Fb-Prophet machine learning model. Applied Computing and Informatics.

Final Report

ORIGINALITY REPORT

25%	18%	20%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Venkatbharat Poleneni, Jahnavi K Rao, Syed Afshana Hidayathulla. "COVID-19 Prediction using ARIMA Model", 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021 Publication	7%
2	www.ijmh.org Internet Source	2%
3	Submitted to Daffodil International University Student Paper	2%
4	www.ncbi.nlm.nih.gov Internet Source	2%
5	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
6	www.medrxiv.org Internet Source	1%
7	El-Sayed Atlam, Ashraf Ewis, M.M. Abd El-Raouf, Osama Ghoneim, Ibrahim Gad. "A New Approach in Identifying the Psychological Impact of COVID-19 on University Student's	1%