Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2020 doi:10.20944/preprints202006.0353.v1

Analysis and Forecast of COVID-19 Pandemic in Pakistan

Abdul Bari Malik

Micromerger, Islamabad, Pakistan

Abstract

The COVID-19 infections in Pakistan are spreading at an exponential rate and a point may soon be reached where rigorous prevention measures would need to be adopted. Mathematical models can help define the scale of an epidemic and the rate at which an infection can spread in a community. I used ARIMA Model, Diffusion Model, SIRD Model and Prophet Model to forecast the magnitude of the COVID-19 pandemic in Pakistan and compared the numbers with the reported cases on the national database. Results depicts that Pakistan could hit peak number of infectious cases between June 2020 and July, 2020.

1 Introduction

The first case of COVID-19 emerged in Pakistan on 26th February 2020 in Karachi, the most populous city of Pakistan. The patient zero had a travel history to Iran and immediately quarantined upon testing positive for the virus. However, the patient was followed by hundreds of pilgrims returning from Iran which were likely carrying the virus that ultimately led to the spread of COVID-19 into the community. Since then the infections have been increasing exponentially and without proper intervention the situation may escalate enough to overwhelm the already struggling healthcare system in the country. In this study we aim to model the COVID-19 pandemic in Pakistan, which will indicate the peak infection day, rate of increase of infections per day and a 90-day forecast to assess the final size of the epidemic.

2 Models Used

The accuracy of traditional forecasting largely depends on the availability of data to base its predictions and estimates of uncertainty. In outbreaks of pandemics there is no data at all in the beginning and then limited as time passes, making predictions widely uncertain.

To forecast confirmed cases of COVID-19, we adopt time series forecasting approaches. I am using models from the ARIMA family, SIR family, PROPHET Models and Diffusion Gradient family.

2.1 ARIMA (Autoregressive integrated moving average)

ARIMA, short for **'Autoregressive Integrated Moving Average'** is actually a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values. Any 'non-seasonal' time series that exhibits patterns and is not a random white noise can be modeled with ARIMA models. Since COVID-19 data for Pakistan depicts some **'non-stationarity'** and '**non-seasonal'** pattern in it and reflects a linearity among data points. That's why ARIMA is our first choice.

ARIMA model in words:

- Predicted Yt = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (upto q lags)

If in future our time series will reflect some seasonal pattern, then we would be need to add seasonal terms and it becomes SARIMA, short for 'Seasonal ARIMA'

2.2 Diffusion Process

I am using a model from a paper by Emmanuelle Le Nagard and Alexandre Steyer, that attempts to reflect the social structure of a diffusion process. The model is also sensitive to when we define the origin of time for the epidemic process. Here, I just took the first point of the time series available. I am exploring the 3d parameter space to find a minimum, using Gradient Descent.

Basic Model for Diffusion:

- $Y(t) = \alpha(m - Y(t))$
  - where ($\alpha > 0$) plus the "typical" initial condition $Y(0) = 0$.
  - Y denotes the cumulative number of adopters at time t
  - m the size of the population of potential adopters
  - α the intensity of the spreads

2.3 SIRD Model of Epidemiology for COVID-19

Since the emergence of COVID-19, several mathematical models have been developed to simulate the rate of infection spread, infections per day and the resolution of the epidemic. The SIRD model refers to the number of susceptible, infected, resolved and deceased cases during an epidemic at any given time. The model assumes that susceptible cases (S), infected cases (I), resolved cases (R) and Deceased cases (D) are compartments and each individual of a given population will pass through the susceptible phase then to the infected phase and finally to the recovered phase.

The SIRD model is a steady state model, therefore the population that is analyzed is static i.e. no one is being born or is dying. Additionally, the model assumes that once a person is infected, they are immune to the disease and therefore cannot contract it again. The SIRD model is ideal for modelling the spread of diseases spread through person to person contact.

Basic SIRD Model:

- Number of people who are (stocks):
  - S_t = Susceptible
  - I_t = Infectious

   - R_t = Resolving
   - D_t = Dead
   - C_t = ReCovered
- Constant population size is N
   - $S_t + I_t + R_t + D_t + C_t = N$
- Susceptible get infected at rate $\beta_t I_t/N$
   - New infections = $\beta_t I_t/N \cdot S_t$
- Infectiousness resolve at Poisson rate γ, so the average number of days that a person is infectious is $1/\gamma$ so $\gamma = .2 \Rightarrow 5$ days
- Post-infectious cases then resolve at Poisson rate θ. E.g. $\theta = .1 \Rightarrow 10$ days
- Resolution happens in one of two ways:
   - Death: fraction δ
   - Recovery: fraction $1 - \delta$
- SIRD Model - Laws of Motion
   - $\Delta St{+}1 = -\beta_t S_t I_t / N$ (new infections)
   - $\Delta It{+}1 = \beta_t S_t I_t/N$ (new infections) $- \gamma I_t$ (resolving infectious)
   - $\Delta Rt{+}1 = \gamma I_t$ (resolving infectious) $- \theta R_t$ (cases that resolve)
   - $\Delta Dt{+}1 = \delta\theta R_t$ (die)
   - $\Delta Ct{+}1 = (1 - \delta)\theta R_t$ (recovered)
   - $D0 = 0$

2.4 Prophet

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well. Prophet automatically evaluates forecast performance and flags issues that warrant manual intervention.

Prophet uses two approaches for saturating forecasts.

- Forecasting Growth
- Saturating Minimum

Prophet uses a linear model for its forecast. When forecasting growth, there is usually some maximum achievable point: total market size, total population size, etc. This is called the carrying capacity, and the forecast should saturate at this point. Prophet also allows you to make forecasts using a logistic growth trend model, with a specified carrying capacity.

For COVID-19 Pakistan forecast, we are using Linear Model. But I have space of implementing a logistic growth function and lag parameter for obtaining Logistic growth with Prophet.

Basic Prophet Model

- Prophet is an additive model with the following components:
   - $y(t) = g(t) + s(t) + h(t) + \epsilon_t$

- g(t) models trend, which describes long-term increase or decrease in the data. Prophet incorporates two trend models, a saturating growth model and a piecewise linear model, depending on the type of forecasting problem.
- s(t) models seasonality with Fourier series, which describes how data is affected by seasonal factors such as the time of the year (e.g. more searches for eggnog during the winter holidays)
- h(t) models the effects of holidays or large events that impact business time series (e.g. new product launch, Black Friday, Superbowl, etc.)
- ϵ_t represents an irreducible error term

3 Data

The data was extracted from daily situation reports published on the National Institute of Health (NIH) Pakistan for a period of 120 days, (from February26, 2020 to June 20, 2020) and was corroborated with the simulation results

4 Analysis

The NIH Pakistan data for the cumulative cases and daily reported cases was plotted to observe the increase in number of cases for a period of 116 days.

Figure (4.1) shows total count of Confirmed Cases, Fatalities and Recovered Cases at provincial level since the pandemic started in Pakistan. The current trajectory suggests an exponential increase in the number of cases since the pandemic started in Pakistan on 26th February 2020 (Figure 4.2). The data on the number of daily reported cases also reflects a general increasing trend as testing for COVID-19 gathers pace (Figure 4.3).

Figures (4.4), (4.5) and (4.6) depicts daily count of last 30 days for confirmed cases, fatalities and recovered cases respectively. While Figures (4.7) and (4.8) reflects weekly increase in number of confirmed cases and fatalities in Pakistan. Figure (4.9) depicts monthly increase in number of confirmed cases. Fatalities and recovered cases in Pakistan.

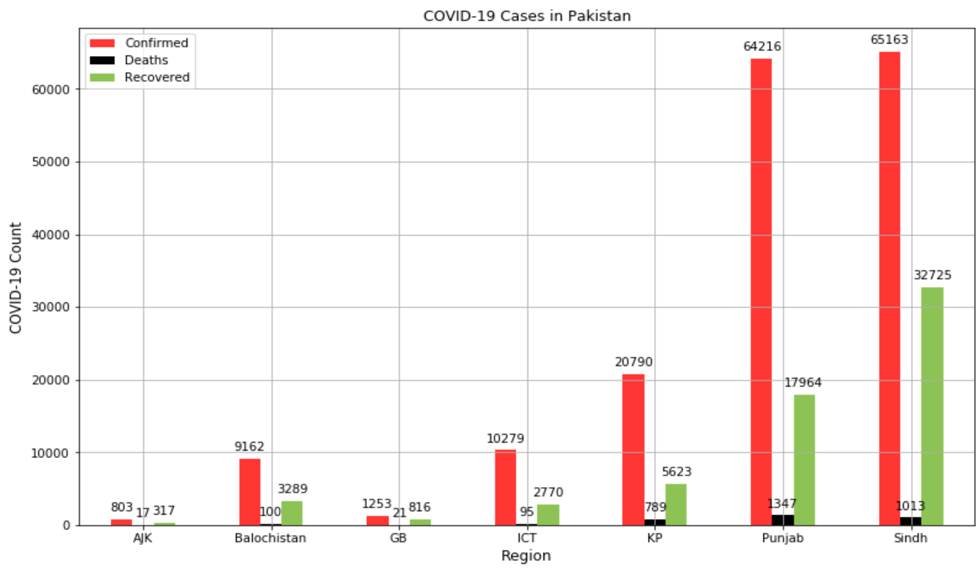

Figure 4-1: The COVID-19 cases in Pakistan since the outbreak began on 26th February 2020

 doi:10.20944/preprints202006.0353.v1

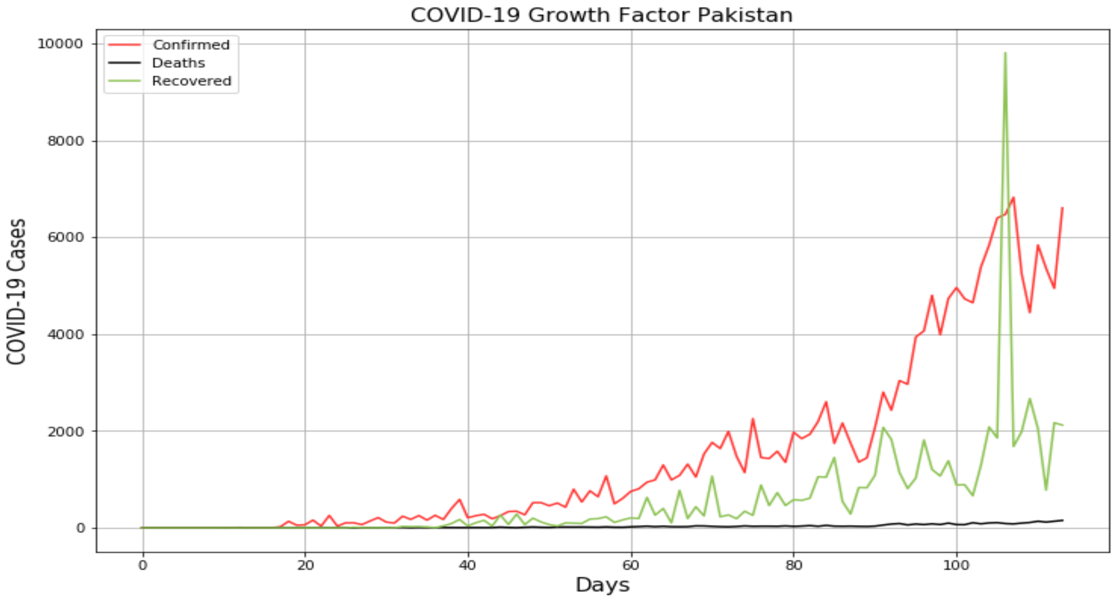

Figure 4-2: The graph shows an exponential growth pattern for the 116-day data extracted from the NIH, Pakistan database. *The x-axis corresponds to the days, whereas the y-axis corresponds to the number of cases.*

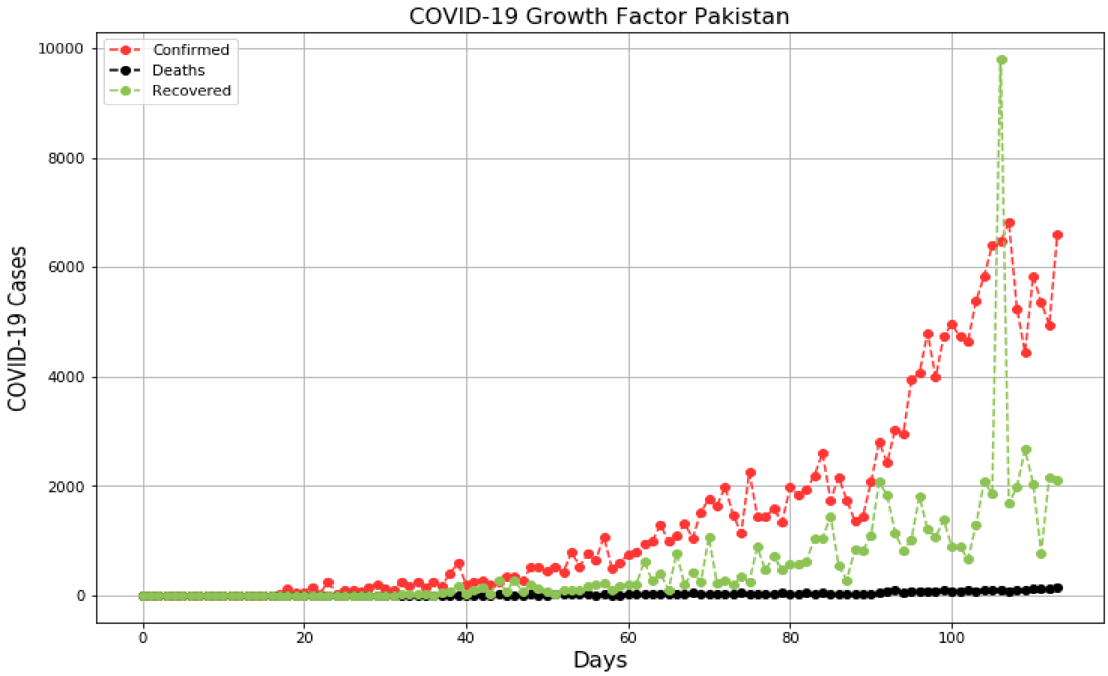

Figure 4-3: Daily reported COVID-19 cases in Pakistan since the outbreak on 26th February 2020. *The graph shows an exponential growth pattern for the -day data extracted from the NIH, Pakistan database. On day 1 of the epidemic the number of cases were 2 and on day 116 the number of daily reported cases were 6604. These numbers may represent the gap in the unreported or asymptomatic cases in Pakistan. The x-axis corresponds to the days, whereas the y-axis corresponds to the number of cases.*

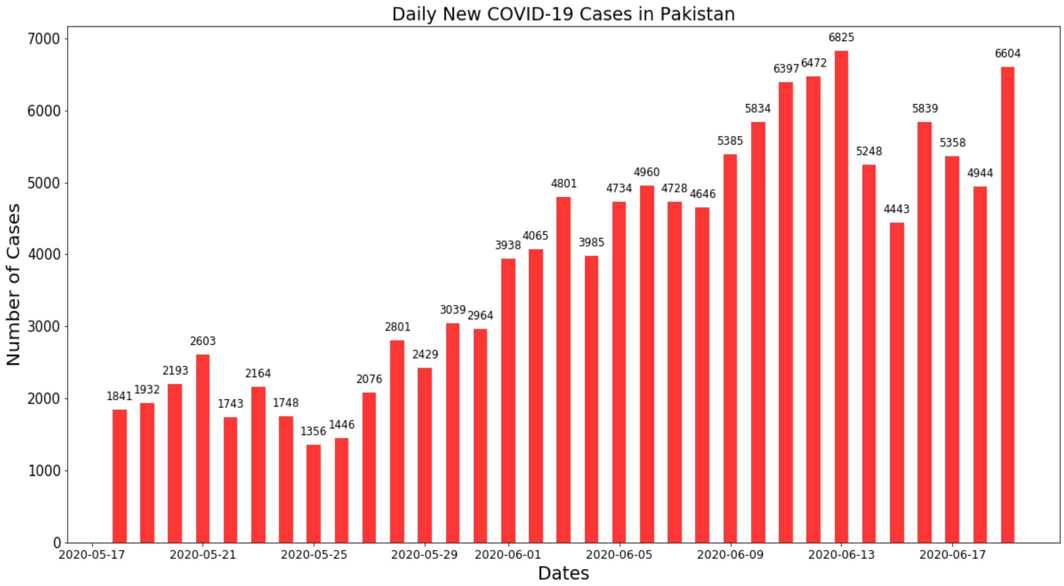

Figure 4-4: Daily Reported Cumulative Tests Positive (Daily Confirmed Cases) in Pakistan

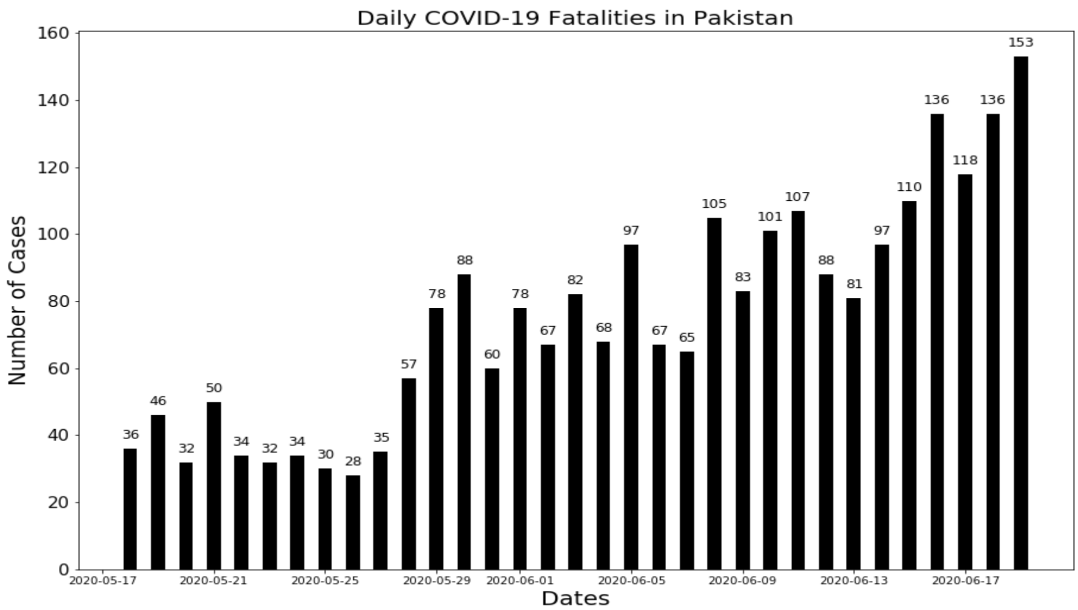

Figure 4-5: Daily reported Cumulative COVID-19 Fatalities in Pakistan

 doi:10.20944/preprints202006.0353.v1

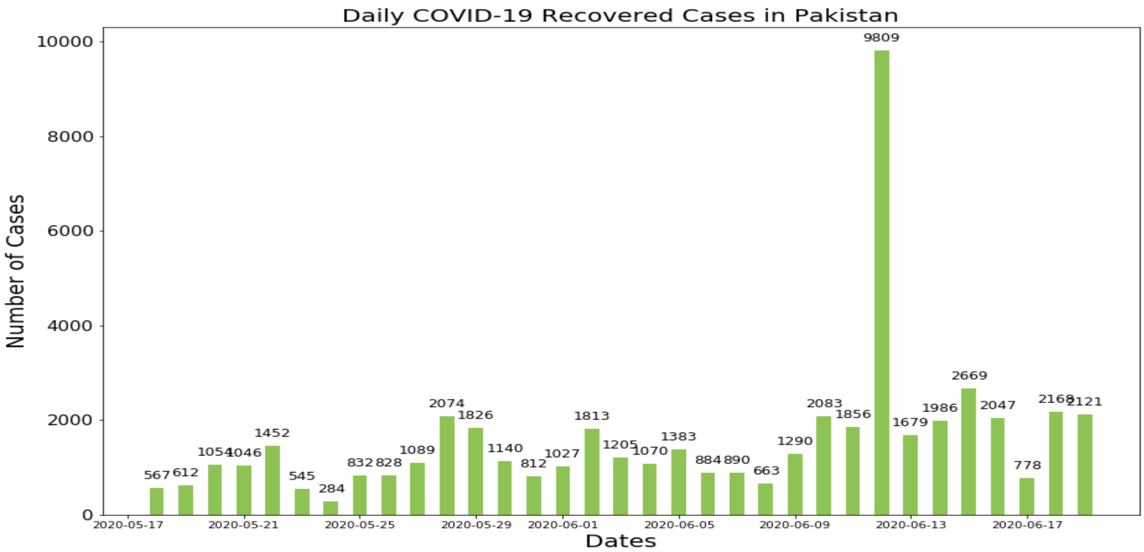

Figure 4-6: Daily Reported Recovered Cases in Pakistan

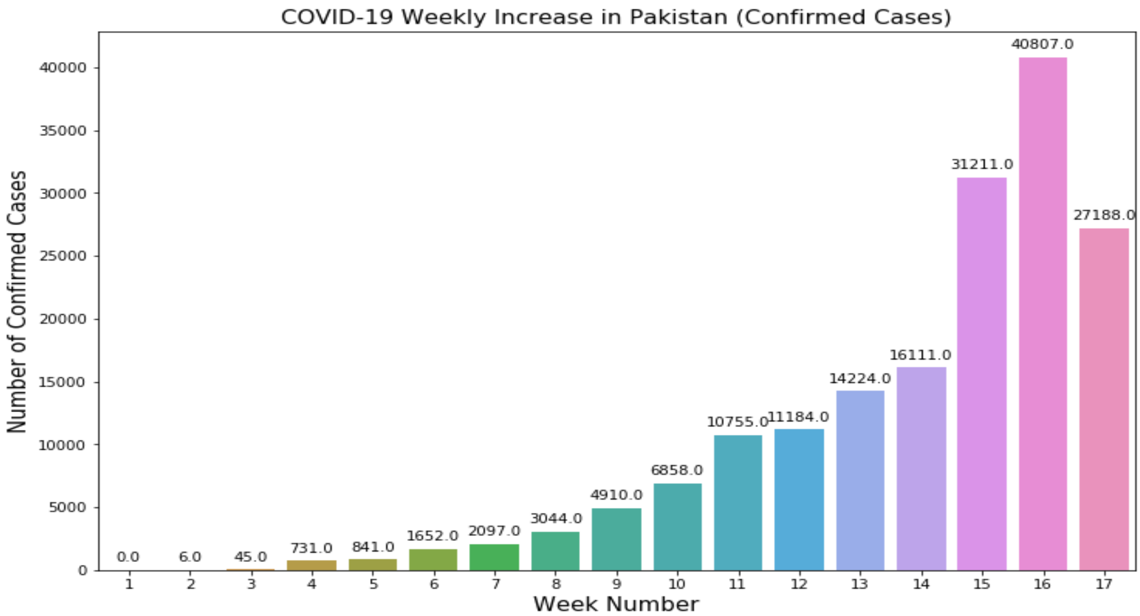

Figure 4-7: Weekly increase in number of Confirmed Cases in Pakistan. The figure shows reported Cumulative Tests Positive (Daily Confirmed Cases) per week since the outbreak began on 26th February 2020.

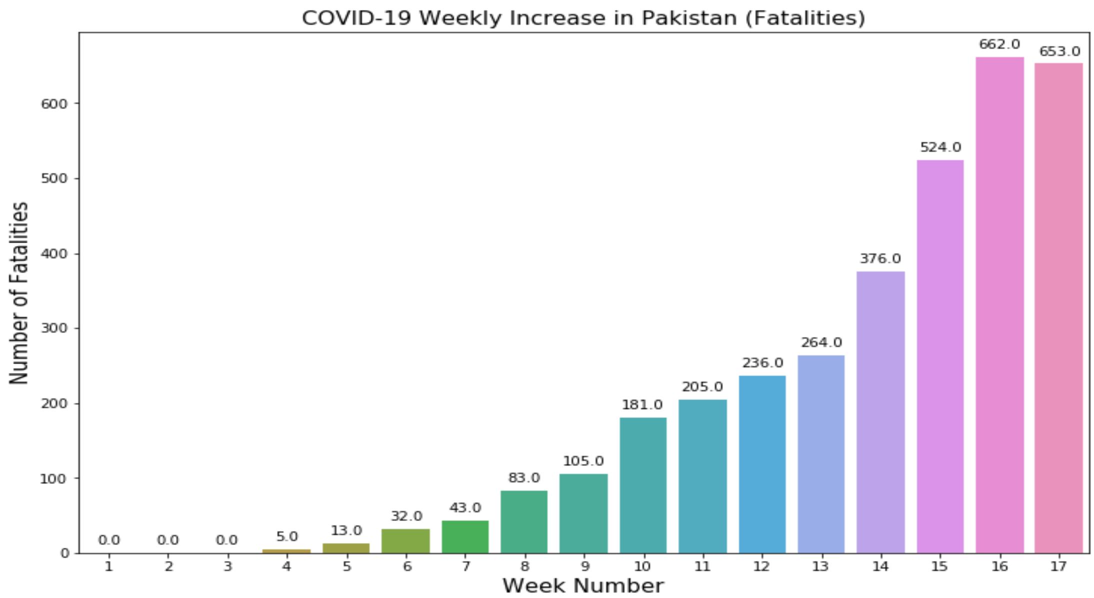

Figure 4-8: Weekly increase in number of Fatalities in Pakistan. The Figure shows increase in number of fatalities per week since the outbreak began on 26th February 2020.

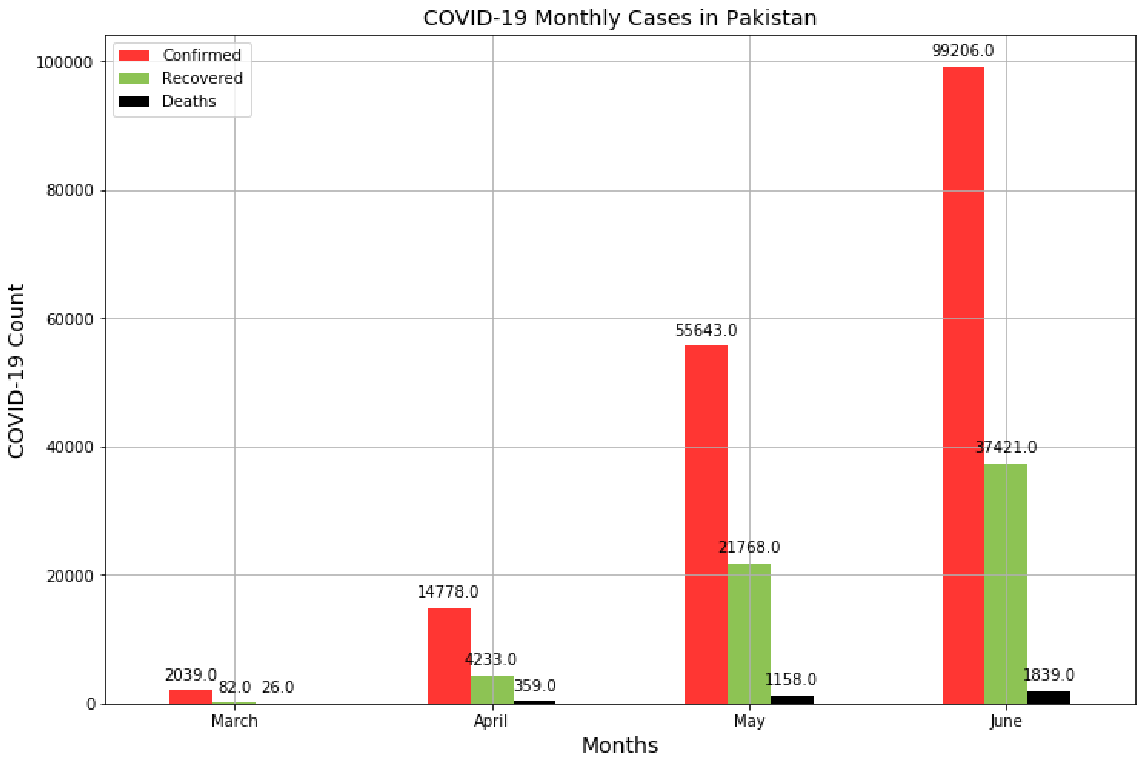

Figure 4-9: Cumulative Monthly COVID-19 cases in Pakistan. The Figure depicts monthly increase in number of positive test cases, fatalities and recovered personnel in Pakistan

5 Projections for PAKISTAN

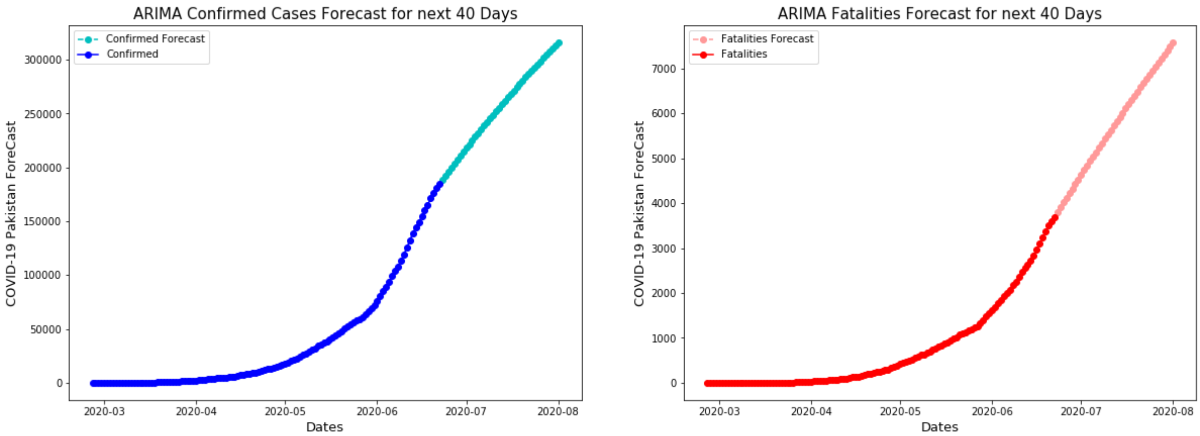

Figure 5-1: COVID-19 estimation using ARIMA. *COVID-19 forecast depicts that COVID-19 Confirmed cases numbers will rise to the count of 300,000 and deceased count will rise to the number 7000 by August 2020.*

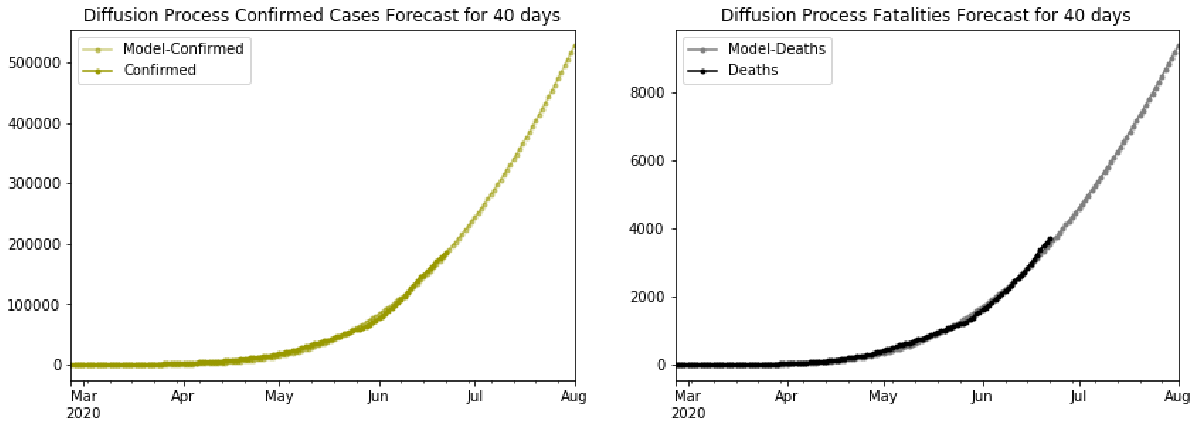

Figure 5-2: COVID-19 Projections using Diffusion Process. *COVID-19 forecast illustrates that COVID-19 Confirmed cases numbers will rise to the count of 500,000 and deceased count will rise above number 8000 by August 2020.*

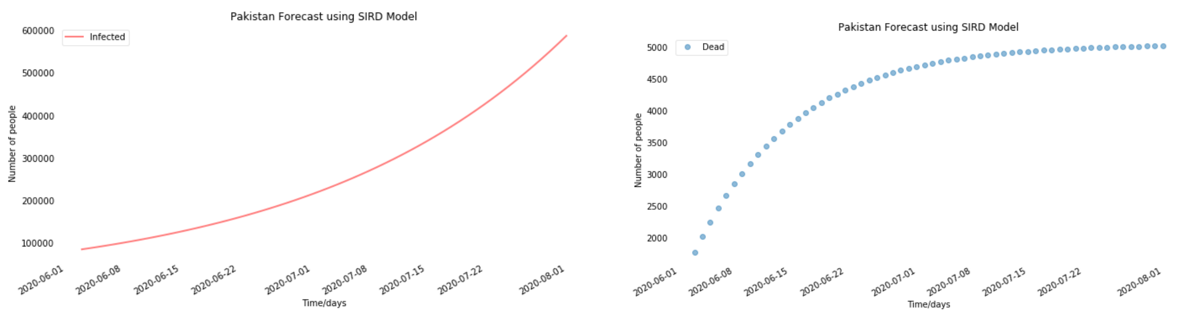

Figure 5-3: COVID-19 Projections using SIRD. *COVID-19 forecast shows that COVID-19 Confirmed cases numbers will rise to the count of 600,000 and deceased tally will rise above figures of 5000 by August 2020.*

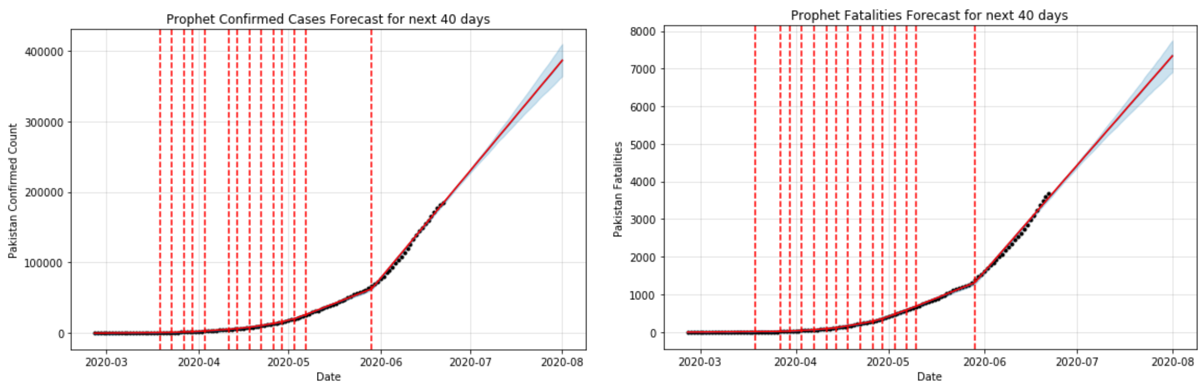

Figure 5-4: COVID-19 Projections using Prophet. *COVID-19 forecast depicts that COVID-19 Confirmed cases numbers will rise to the tally of 600,000 and deceased count will rise above figures of 5000 by August 2020.*

	Projection_Dates	ARIMA_Confirmed	ARIMA_Deaths	Diffusion-Confirmed	Diffusion-Deaths	Prophet_Confirmed	Prophet_Deaths	SARIMAX_Confirmed	SARIMAX_Deaths
0	2020-06-23	188856	3800	190569	3660	190324	3699	189224	3799
1	2020-06-24	192674	3906	196823	3771	195528	3793	193407	3900
2	2020-06-25	196397	4012	203201	3884	200731	3890	197557	4001
3	2020-06-26	200101	4116	209705	3999	206247	3993	201720	4102
4	2020-06-27	203780	4220	216335	4116	211280	4088	205883	4202
5	2020-06-28	207416	4323	223092	4234	216295	4182	210044	4301
6	2020-06-29	211029	4426	229974	4355	221364	4281	214206	4400
7	2020-06-30	214612	4528	236984	4478	226383	4369	218368	4499
8	2020-07-01	218163	4630	244120	4602	232106	4470	222530	4597
9	2020-07-02	221688	4732	251384	4729	237356	4576	226691	4694
10	2020-07-03	225184	4833	258775	4857	243103	4681	230853	4791
11	2020-07-04	228652	4933	266293	4988	248592	4784	235015	4887
12	2020-07-05	232093	5033	273939	5120	253540	4880	239177	4983
13	2020-07-06	235507	5133	281713	5254	259087	4982	243338	5078
14	2020-07-07	238895	5232	289613	5390	264538	5075	247500	5172
15	2020-07-08	242257	5331	297642	5528	270052	5188	251662	5266
16	2020-07-09	245593	5429	305797	5668	276133	5287	255824	5360
17	2020-07-10	248904	5527	314080	5809	281896	5408	259985	5453
18	2020-07-11	252190	5624	322491	5953	287498	5500	264147	5545
19	2020-07-12	255452	5721	331028	6098	293023	5607	268309	5637
20	2020-07-13	258689	5818	339692	6246	298588	5717	272471	5729
21	2020-07-14	261904	5914	348482	6395	303700	5812	276633	5820
22	2020-07-15	265095	6009	357399	6546	309741	5922	280794	5910
23	2020-07-16	268263	6105	366442	6698	315722	6031	284956	6000
24	2020-07-17	271408	6200	375611	6853	321650	6153	289118	6089
25	2020-07-18	274532	6294	384905	7009	327257	6254	293280	6178
26	2020-07-19	277634	6388	394324	7167	333012	6358	297441	6267
27	2020-07-20	280714	6482	403867	7327	338298	6471	301603	6354
28	2020-07-21	283773	6575	413535	7489	343994	6563	305765	6442
29	2020-07-22	286811	6668	423326	7652	350361	6683	309927	6529
30	2020-07-23	289829	6760	433240	7817	356379	6801	314088	6615
31	2020-07-24	292827	6852	443277	7984	361983	6908	318250	6701
32	2020-07-25	295805	6944	453436	8152	367760	7019	322412	6786
33	2020-07-26	298764	7035	463717	8323	372810	7116	326574	6871
34	2020-07-27	301704	7126	474118	8494	379145	7234	330735	6956
35	2020-07-28	304624	7216	484639	8668	384883	7341	334897	7040
36	2020-07-29	307526	7306	495280	8843	390759	7452	339059	7123
37	2020-07-30	310410	7396	506040	9020	396548	7574	343221	7206
38	2020-07-31	313276	7486	516919	9198	402284	7677	347382	7289
39	2020-08-01	316124	7575	527914	9378	408250	7795	351544	7371

Figure 5-5: COVID-19 Projections for Pakistan: Table shows 40 days projections using ARIMA family, Diffusion and Prophet.

6 Conclusion

As the epidemiological data suggests, a huge number of population is susceptible to be affected by COVID-19 infection. The government has already closed the educational institutions and the partial lockdown has been imposed in the country. Efforts should be made to reduce the spread of the virus through social distancing, contact tracing, isolation and quarantine measures. Most importantly, massive testing should be done in the potential population to point out the silent spreader of the virus to reduce the disease spread. There should be a strong communication along with data and resource sharing between central and provincial authorities. Our frontline army is now our healthcare professionals. Steps should be taken to reduce the risk of healthcare providers being infected from the patient by providing recommended personal protective equipment (PPE).

7 References

1. Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities by Jesus Fern ´ andez-Villaverde Charles I. Jones.
2. Hethcote HW. Mathematics of infectious diseases. SIAM Review. 2000;42(4):599-653. doi:10.1137/S0036144500371907
3. Acemoglu, Daron, Victor Chernozhukov, Ivan Werning, and Michael D. Whinston, “A ´ Multi-Risk SIR Model with Optimally Targeted Lockdown,” Technical Report, MIT 2020.
4. Brauer, F., C. Castillo-Chavez, and Z. Feng, Mathematical Models in Epidemiology, Springer New York, 2019.
5. Imperial College COVID-19 Response Team, “Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand,” Technical Report,
6. Social structure of a diffusion process model from a paper by Emmanuelle Le Nagard and Alexandre Steyer.
7. Imperial College 2020. Johns Hopkins University CSSE, “2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository,” 2020. Center for Systems Science and Engineering.