Forecasting COVID-19 Patients & Power Demand in Bangladesh

by

Saad Ebna Azad (160021122)
Atif Mohd. Anwar (160021150)
Fahim Md. Saharear (160021156)

BACHELOR OF SCIENCE
IN
ELECTRICAL AND ELECTRONIC ENGINEERING

Department of Electrical and Electronic Engineering
Islamic University of Technology (IUT)
Board Bazar, Gazipur-1704, Bangladesh.
March, 2021.

CERTIFICATE OF APPROVAL

The thesis titled, “**Forecasting COVID-19 Patients & Power Demand in Bangladesh**” accepted as partial fulfillment of the requirement for the Degree BACHELOR OF SCIENCEIN ELECTRICAL AND ELECTRONIC ENGINEERING of Islamic University of Technology (IUT).

Approved By:

Dr. Khondokar Habibul Kabir
(Supervisor)
Professor,
Electrical and Electronic Engineering Department,
Islamic University of Technology (IUT), Gazipur.

Declaration of Candidate

It is hereby declared that this thesis report is only submitted to The Electrical Engineering Department. Any part of it has not been submitted elsewhere for the award of any Degree or Diploma.

Saad Ebna Azad
Student Id.: 160021122
Date:

Atif Mohd. Anwar
Student Id.: 160021150
Date:

Fahim Md. Saharear
Student Id.: 160021156
Date:

Dedicated to our beloved parents & respected supervisor Professor Dr. Khondokar Habibul Kabir for always guiding us to the right path.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
CSSE	Center for Systems Science and Engineering
IEDCR	Institute of Epidemiology Disease Control And Research
ML	Machine Learning
RMSE	Root Mean Squared Error
SEIR	Susceptible Exposed Infectious Removed
SVM	Support Vector Machine
ARIMA	Auto Regressive Integrated Moving Average
AR	Auto Regressive
MA	Moving Average

Acknowledgment

All praise and gratitude be to Allah, the most beneficent, the most merciful.

Abstract

This research gives a better understanding & analysis of the global pandemic situation of COVID-19 in Bangladesh. Using machine learning this accurately forecasts new cases, deaths & recoveries. Then it forecasts the necessary hospital seats & power demands for those seats. This hopefully will provide a better support for the struggling power sector of developing countries like Bangladesh. This will also help to prepare other countries who are struggling for such epidemic situations in the future.

Chapter 1

Introduction

We aimed to systematically review and critically appraise all currently available prediction models for covid-19, in particular models to predict the risk of covid-19 infection or being admitted to hospital with the disease, models to predict the presence of covid-19 in patients with suspected infection, and models to predict the prognosis or course of infection in patients with covid-19. [1] To find the global impact of the novel coronavirus (COVID-19), we need accurate forecasting the spread of confirmed cases as well as analysis of the number of deaths and recoveries. In outbreaks of epidemics there is no data at all in the beginning and then limited as time passes, making predictions widely uncertain. [2] We developed a modified stacked auto-encoder for modeling the transmission dynamics of the epidemics & applied this model to real-time forecasting the confirmed cases of Covid-19. [3] R_0 is the transmission rate given that the population has no immunity from past exposures or vaccination, nor any deliberate intervention in disease transmission. The number of infections grow and spread in the population if R_0 > 1. They have found that R_0 from India ($R_0 \approx 0.43$) is much smaller than the rest of the world (1.5 < R_0 < 2.5), and the numbers reported in India may not be reflective of the actual number of cases. [4] One month ahead predictions for 3 different scenarios (R_0 = 1.5, R_0 = 2.25, R_0 = 3), estimates the daily number of COVID-19 cases, hospitalizations and deaths, the needs in ICU beds per Region and the reaching date of ICU capacity limits. [5] For modified SEIIR model into a SEII$_s$R model where we consider the isolation process of the infected, R_0 = 3 is found in the model. [6] The ARIMA (1,1,0) model was based upon the parameter test and Box–Ljung test provides a ten-day forecast, which shows a steep upward trend of the spread of the COVID-19. [7] Despite all the problems that make the prediction of cases of COVID-19 a challenging task, the Holt's model can be an adequate alternative to the traditional S-shaped curves if their assumptions are adequately verified and validated by experts. [8] Applying Holt's second-order exponential smoothing method and autoregressive integrated moving average (ARIMA) model, we generate 10-day ahead forecasts of the likely number of infected cases and deaths COVID-19. [9] The time series modeling of this study found that the COVID-19 new cases data have fitted the ARIMA (0,1,0), ARIMA (0,1,0), and ARIMA (0,1,14) models, respectively. [10] Facebook's prophet model had some overestimation & underestimation of the daily cases, the linear model in actual vs predicted cases gave a p-value much lower with a R_2 value of 0.8919. [11] The ensemble provides updated COVID-19 forecasts every 24 hours, combines an Auto-Regressive Integrated Moving Average model (ARIMA), Prophet - an additive regression model developed by Facebook, and a Holt-Winters Exponential Smoothing model combined with Generalized Autoregressive Conditional Heteroscedasticity (GARCH). [12]

COVID-19 pandemic is part of the worldwide pandemic of coronavirus disease 2019 caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Wuhan, Hubei, China is the birthplace of this catastrophic virus. Since then, it has travelled all over the world reaching Bangladesh at March 8,2020. It has grown steadily ever since and is still on the rise.

Machine Learning studies algorithms that allow machines recognizing patterns, construct prediction models or generate images or videos through learning. [13] [14] ML is the study of computer algorithms that improve automatically through experience. It is a part of artificial intelligence. ML algorithms build a model based on sample data, which are called 'training data'. All these in order to make predictions or decisions without being explicitly programmed to do so. ML algorithms are used in a wide variety of applications where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks. For simple tasks assigned to computers, it is possible to program algorithms telling the machine how to execute all steps required to solve the problem at hand; on the computer's part, no learning is needed. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine develop its own algorithm, rather than having human programmers specify every needed step.

Smart Grid is important as it will take us towards energy independence and environmentally sustainable economic growth. [15] Smart Grid is an advanced digital two-way power flow power system capable of self-healing, adaptive, resilient and sustainable with foresight for prediction under different uncertainties. [16] Traditional energy grid designs marginalize the value of information and energy storage, but a truly dynamic power grid requires both. [17] It is a data communications network which is integrated with the power grid to collect and analyze data that are acquired from transmission lines, distribution substations, and consumers. [18] Includes- Advanced metering infrastructures (of which smart meters are a generic name for any utility side device even if it is more capable e.g., a fiber optic router). Smart distribution boards

& circuit breakers integrated with home control & demand response (behind the meter from utility perspective). Load control switches and smart appliances, often financed by efficiency gains on municipal programs (e.g., Property Assessed Clean Energy). Renewable energy resources, including capacity to charge parked (electric vehicle) batteries or larger arrays of batteries recycled from these, or other energy storage.

1.1 Problem Statement

This study aims to make us better understand about pandemic situations & future epidemics like COVID-19. This will make a prediction of total cases like- affected, deaths, recoveries & then use them to make an educated guess of necessary power demand in order to treat all of our patients effectively via hospital ICU units or whatever that is necessary.

Research Question- How much power we need to provide to the hospitals if we were to accumulate all the COVID-19 patients?

1.1.1 Research Gap

There have been lots of research papers incorporating predictions of COVID-19, treatments of COVID-19, vaccines of COVID-19. And also there have been research papers about the power demand which is needed for our country. But in this present scenario, we need extra amounts of electricity & power in our hospitals for treating the COVID-19 patients properly. Unfortunately there has no combined research paper of COVID-19 predictions and extra power which is needed for hospitals in this pandemic situation.

1.1.2 Problem Identification

No interaction between the current exploration paper of COVID-19 expectations and Power system. Lack of hospital bed and power which is required for hospitals. Set up the hospital power system for recuperating this kind of pandemic circumstance sooner rather than later.

1.1.3 Motivation

Lots of people died due to COVID-19 & many of them did not get proper treatment (lack of ventilators & ICU beds) in the hospital. To fulfill this shortage of hospital bed, we need to prepare our hospitals with sufficient number of beds. Those extra numbers of hospital beds require a certain amount of electrical power. So that we need to equip our national grid system to mitigate this extra power which is needed to provide proper treatment for the COVID-19 patients in this pandemic situation.

1.1.4 Scope of the Research

This research mainly focuses on current situations of COVID-19 & it's impact on Bangladesh. While also creating a reliable ML algorithm model to make accurate forecast of COVID-19 affected cases & the power demand to occupy all those active patients in our hospitals.
In this research the effect of temperature, humidity etc. is not taken into consideration. If this parameter are changed the results obtained in this research may also vary . Despite the research being based on the COVID-19 in Bangladesh, the algorithms in the model & the power demand can also be applicable to any region given enough data.

1.2 Research Objectives

- To forecast & analyze COVID-19 scenario in Bangladesh.
- To find out the extra amount of bed needed & total power demand for those requirements.
- To implement a smart grid network system to solve our existing power demand issues.

1.3 Research Outcome

- Forecast of COVID-19 cases in Bangladesh.
- Forecast of extra amount of bed & power needed in hospitals.
- Proposal of a smart grid network system to meet this extra power demand.

1.4 Novelty of the Research

Despite being a relatively new topic, there has been a lot of research done on COVID-19 due to the impact it had on the whole world. People have been affected one way or another somehow by COVID-19. So it is no shocker to see all the research being done or still being on the works as it deserves all the attention it is getting in order to fully nullify COVID-19 & prevent such pandemic situations in the upcoming days.
Like other researches this research also uses ML algorithms to predict the future COVID-19 situations. This research deals with different algorithm combinations but the main

idea remains same, to accurately forecast COVID-19 cases for upcoming days. But this research also deals with the power demands associated with the increase of COVID-19. As mentioned earlier, this part of research has previously been ignored for some reason as most people solely dedicated their research to make accurate future forecasts for COVID-19 cases. Though power demand can not be ignored as we need power to treat our ever-increasing COVID-19 patients that we are trying so hard to predict through all these research. So this research will not only help us to see & analyse the worst outcomes of COVID-19 but also prepare us to tackle the situation with necessary power required.

1.5 Overview of the Methodology

This research is mainly divided into two parts.
In the 1st part, we separate the data of Bangladesh from the rest of the world to analyse & inspect the current scenario of COVID-19. Then various ML algorithms proposed in this research is implemented to forecast the 3 separate cases- affected, dead & recovered cases.
In the 2nd part, all the predicted cases are then used to forecast the ICU & non-ICU hospital beds required to treat all the predicted patients by COVID-19. This can then be implemented to figure out the power demand for all these hospital beds.

1.6 Organizations of the Thesis

The rest of the paper is designed as follows:
Chapter 2 presents the working procedure of algorithms
Chapter 3 describes some related works in the literature
Chapter 4 presents a working diagram & how the algorithms work
Chapter 5 analyzes the experiment results
Chapter 6 states about the discussion
Chapter 7 presents conclusions & future work

Chapter 2

Background

This chapter discusses previous researches that are related to ensemble architectures, learning algorithms and argumentation approaches.

Related Works

In the research [2] they wanted to know how the global pandemic COVID-19 is affecting the world & to do so they need accurate forecasting of the number of affected people, recovered cases, deaths. Which requires a huge amount of data & that data must be reliable.
However, the main caveat is that future does not repeat itself. Assuming that it does, their forecast suggests the continuous rise of confirmed cases of COVID-19 & it's growth if proper measures are not taken in due time.

As stated before, accuracy of a forecast depends on data. But for epidemics & pandemics there is no data in the beginning and very limited as time passes. Moreover, the data is not 100% accurate due misreports and a new category "Clinically Diagnose" was added to "Lab Diagnosed"; which created more confusion. Predicting the future of epidemics and pandemics is much more difficult as the number of cases to be studied can be measured in one hand.
All necessary data were taken from CSSSE at John Hopkins University at GitHub.com which included both "Clinically Diagnose" and "Lab Diagnosed". They used models from exponential smoothing family & followed a pragmatic approach assuming trend will continue in the future.

It spans over 5 rounds.

- 1st round: 01/02/20 till 10/02/20 (absolute percentage error of 388%)
- 2nd round: 11/02/20 till 20/02/20 (absolute percentage error of 7.7%)
- 3rd round: 21/02/20 till 01/03/20 (absolute percentage error of 6.2%)

- 4th round: 02/03/20 till 11/03/20 (absolute percentage error of 12.1%)
- 5th round: 12/03/20 till 21/03/20 (absolute percentage error of either 50%, 70% or 90%)

Despite the very small percentages of recovered cases until the end of January (less than 2%), currently, about 1 out of 2 confirmed cases have recovered (52.8% of the total confirmed cases). Moreover, the ratio of recovered cases versus deaths is currently above 14:1. After 2nd & 3rd period, they assumed COVID-19 would not go outside of China due to its slowing down trend. But they were wrong.

Research [4] which is; a comparison of COVID-19 between India & several states of US has been made and a basic reproduction number R0 = 1.4-3.9 was determined. Exponential & classic susceptible-infected-recovered (SIR) models are used for short and long-term prediction. Which resulted in India entering equilibrium by the end of May with total 13,000 infected.

COVID-19 is transmitted by inhalation or contact with infected droplets or fomites, and the incubation period may range from 2-14 days. It can happen to anyone but has the most devastating effect on older or people with a medical history. R_0 is the transmission rate given that the population has no immunity from past exposures or vaccination, nor any deliberate intervention in disease transmission. Where it is high for rest of the world (1.5 <R_0 <2.5), it is very small for India ($R_0 \approx 0.43$). Though it may not be the actual representation of no of cases. The uncertainty due to exclusion of asymptomatic cases can be a major limitation in predictions with these models. India is currently on stage 1 & 2 but will soon enter pandemic stage-3 (community transmission) if proper measures are not taken.
Although 'Exponential Model' fit is not good for long term prediction, it helps to see the worst-case scenario. 'SIR Model' is there for long term prediction. The no of affected cases by April 30, according to-

1. Exponential Model – 0.5 million
2. SIR Model – 12,416

The main problem about these models is that the social distancing fact is not taken into consideration. Still at this rate equilibrium will come by the end of May. However, all might change if social distancing is not maintained & India enter stage-3.
The transmission rate & reproduction number R_0 for India is similar to Washington as outbreak in India started after 9 days in Washington. Despite many reports questioning testing strategies of India. Here using SIR model R_0 is 1.504, which is close to the values found by others using ARIMA model. After April 8, no of confirmed cases

may reduce if social distancing is maintained.

Research [19] describes that Nonlinear models are widely used to forecast the spreading of the disease and capture the probability of cases from susceptible to infected. In this study, they have used Auto regression integrated moving average model (ARIMA) and Richard's model in the R-language platform to predict some trajectories associated with COVID-19 pandemic in the coming days in India. Here some statistical phenomenological models are used to detect and analyze the disease-based trajectory model for prediction purposes.

To analyze the aggregate, this research used four models. Which predicted that there would be 5200 new cases (95% CI: 4650 to 6002) through the ARIMA model versus be 6378 (95% CI: 4904 to 7851) through Richard model at the end of April 2020. It then estimated that there would be a total of 197 (95% CI: 118 to 277) deaths and drop down in the recovery rates will reach around 501 (95% CI: 245 to 758) by the end of April 2020.

Time series models provide a different and unique approach to time series forecasting. Basically, for the time series forecast, two approaches are widely used. Exponential smoothing and Time Series Models like ARIMA and Richard's. While exponential smoothing models are based on a description of the trend and seasonality in the data, Time Series, like ARIMA models, aims to describe the autocorrelations in the data.

Research [20] discusses the building electrical energy forecasting method using artificial intelligence(AI) methods such as support vector machine (SVM) & artificial neural networks (ANN). Superfluous energy is needed to drive the global demand; & at the same time, the environment needs to be kept safe. Since the complexity of building energy system is very high due to several factors, the ability of ANN & SVM in performing non-linear analysis is an advantage in executing buildings energy consumption forecasting.

Many researchers are working diligently to improve this field & found out a lot. Mainly enhanced ANN & SVM and experimented with both of them which results in more efficient scenario. It is still being worked on & can offer a lot.

Chapter 3

Proposed Models

This chapter discusses all the algorithms used & how each one of them operate.

Epidemiological Algorithms

3.1 Polynomial Regression

Polynomial Regression models are generally fit utilizing the technique for least squares. The least-squares technique limits the change of the impartial assessors of the coefficients, under the states of the Gauss–Markov hypothesis. The objective of this is to demonstrate the normal estimation of a reliant variable y regarding the estimation of a free factor (or vector of autonomous factors) x. In straightforward direct relapse, the model utilized is

$$Y = \beta_0 + \beta_1 + \epsilon \tag{3.1}$$

where ϵ is a surreptitiously arbitrary mistake with mean zero adapted on a scalar variable x. In this model, for every unit increment in the estimation of x, the contingent desire for y increments by β_1 units. The relapse work is direct regarding the obscure boundaries β_0, β_1, ...

Consequently, for least squares examination, the computational and inferential issues of polynomial relapse can be totally tended to utilizing the methods of numerous relapses.

3.2 SVM Regression

In AI, uphold vector machines (SVMs, additionally uphold vector organizations) are administered learning models with related learning calculations that dissect information utilized for arrangement and relapse investigation. At the point when information is unlabeled, managed learning is unimaginable, and an unaided learning approach is required, which endeavors to discover characteristic bunching of the information to

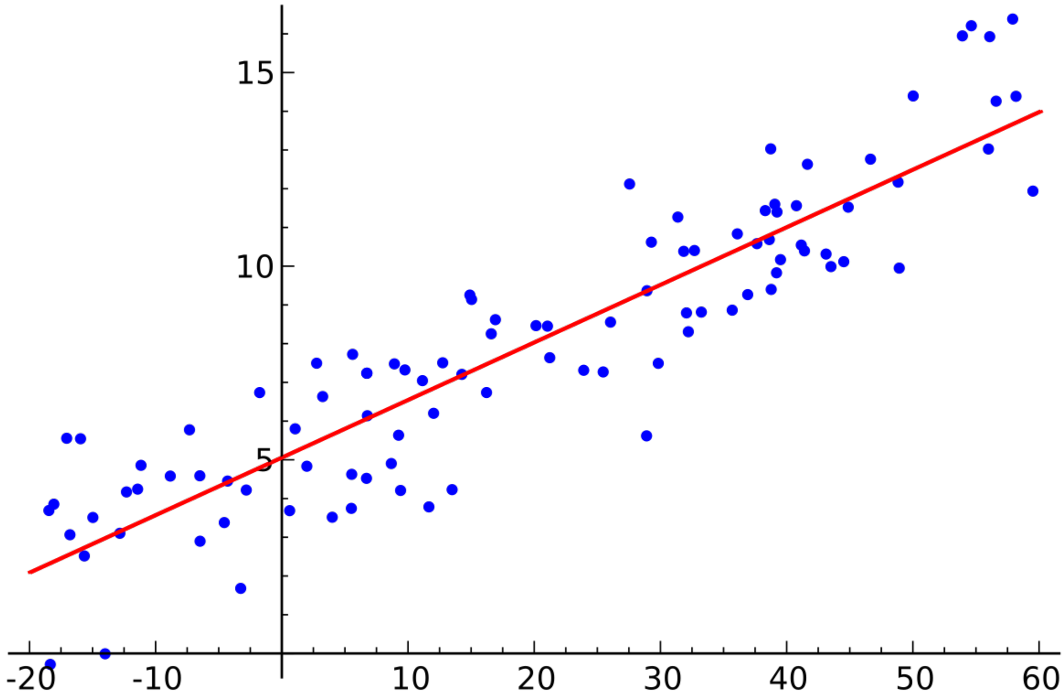


Figure 3.1: Polynomial Regression

gatherings, and afterward map new information to these shaped gatherings. The help vector clustering calculation, applies the insights of help vectors, created in the help vector machines calculation, to arrange unlabeled information, and is one of the most generally utilized bunching calculations in modern applications.

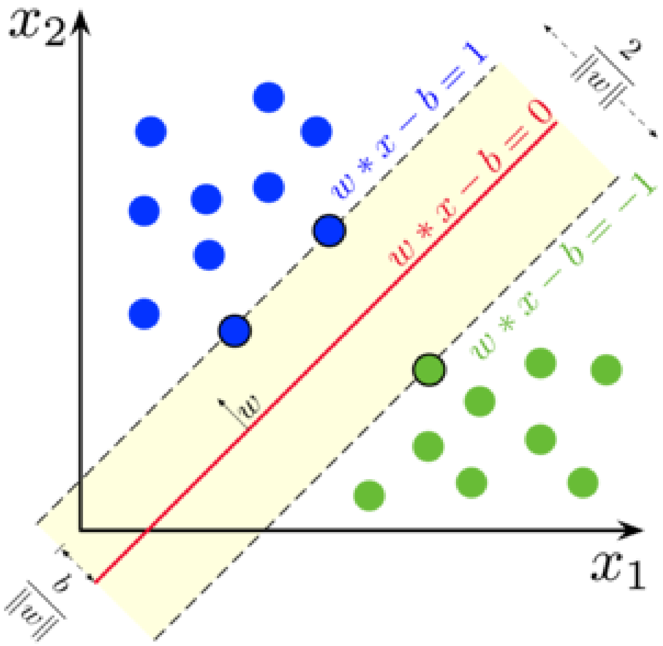


Figure 3.2: SVM Regression

3.3 Holt's Linear Model

Holt's two-parameter model, also called direct exponential smoothing, is a famous smoothing model for anticipating information with pattern. Holt's model has three separate conditions that cooperate to create a last estimate. The first is a fundamental smoothing condition that changes the last smoothed an incentive for last period's pattern. The pattern itself is refreshed after some time during that time condition, where the pattern is communicated as the distinction between the last two smoothed qualities. At last, the third condition is utilized to create the last conjecture.
The model takes the following form for all $i > 1$
$u_1 = y_1 \qquad v_1 = 0$

$$u_1 = \alpha y_i + (1 - \alpha)(u_{i-1} + v_{i-1}) \tag{3.2}$$

$$v_i = \beta(u_i - u_{i-1}) + (1 - \beta)v_{i-1}\hat{y}_{i+1} = u_i + v_i \tag{3.3}$$

where,

$0 < \alpha \leq 1 \qquad \& \qquad 0 \leq \beta \leq 1$

An alternative form of these equations is
$u_1 = y_1 \qquad v_1 = 0$

$$u_1 = u_{i-1} + v_{i-1} + \alpha e_i \tag{3.4}$$

$$v_1 = v_{i-1} + \alpha\beta e_i \tag{3.5}$$

$$\hat{y}_{i+1} = u_i + v_i \tag{3.6}$$

Here,

$$e_1 = y_i - (u_{i-1} + v_{i-1}) = y_1 - \hat{y}_{i-1} \tag{3.7}$$

Note that if $\beta = 0$, then the Holt model is equivalent to the Single Exponential Smoothing model.

3.4 Holt's Winter Model

Holt and Winters extended Holt's method to capture seasonality. The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations — one for the level l_t,

one for the trend b_t, &
one for the seasonal component s_t,
with corresponding smoothing parameters α, β^* and γ. We use mm to denote the frequency of the seasonality, i.e., the number of seasons in a year.
For example, for quarterly data m=4, and for monthly data m=12.
The component form for the additive method is:

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t+h-m(k+1)} \\
\ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},
\end{aligned}$$

The component form for the multiplicative method is:

$$\begin{aligned}
\hat{y}_{t+h|t} &= (\ell_t + hb_t)s_{t+h-m(k+1)} \\
\ell_t &= \alpha\frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma\frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}
\end{aligned}$$

3.5 ARIMA Model

ARIMA, short for 'Auto Regressive 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.
When two out of the three terms are zeros, the model may be referred to base on the non-zero parameter, dropping "AR", "I" or "MA" from the acronym describing the model.
For example, ARIMA (1,0,0) is AR (1), ARIMA (0,1,0) is I (1) & ARIMA (0,0,1) is MA (1)
An ARIMA model is characterized by 3 terms: p, d, q.
Here,
p is the order of the AR term
q is the order of the MA term

d is the number of differencing required to make the time series stationary

3.5.1 AR (Auto Regressive) Model

The autoregressive model indicates that the yield variable relies directly upon its own past qualities and on a stochastic term (a defectively unsurprising term); in this manner the model is as a stochastic distinction condition (or repeat connection which ought not be mistaken for differential condition). AR model is defined as

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t \tag{3.8}$$

$$X_t = c + \sum_{i=1}^{p} \phi_i B^i X_{t-i} + \epsilon_t \tag{3.9}$$

$$\phi[B] X_t = c + \epsilon_t \tag{3.10}$$

3.5.2 MA (Moving Average) Model

Instead of utilizing the past estimations of the conjecture variable in a relapse like model.

$$Y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_a \epsilon_{t-a} \tag{3.11}$$

Where ϵ_t is repetitive sound. We allude to this as a MA (a) model, a moving normal model of request a. obviously, we don't watch the estimations of ϵ_t so it isn't generally a relapse in the typical sense.

3.6 Facebook's Prophet Model

Prophet is a system for determining time arrangement information dependent on an added substance model where non-direct patterns are fit with yearly, week by week, and every day irregularity, in addition to occasion impacts. It works best with time arrangement that have solid occasional impacts and a few periods of recorded information. Prophet is vigorous to missing information and movements in the pattern, and normally handles anomalies well. Although it is Facebook's Core Data Science team, it is an open-source software.
We use a decomposable time series model with three main model components: trend,

seasonality, and holidays. They are combined in the following equation:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{3.12}$$

$g(t)$: piece-wise linear or logistic growth curve for modelling non-periodic changes in time series
$s(t)$: periodic changes (e.g., weekly/yearly seasonality)
$h(t)$: effects of holidays (user provided) with irregular schedules
ϵ_t : error term accounts for any unusual changes not accommodated by the model

Chapter 4

Methodology

This chapter introduces a pipeline architecture that integrates all of our algorithms. It also introduces a formal flowchart of the proposed system.

4.1 Working Algorithms

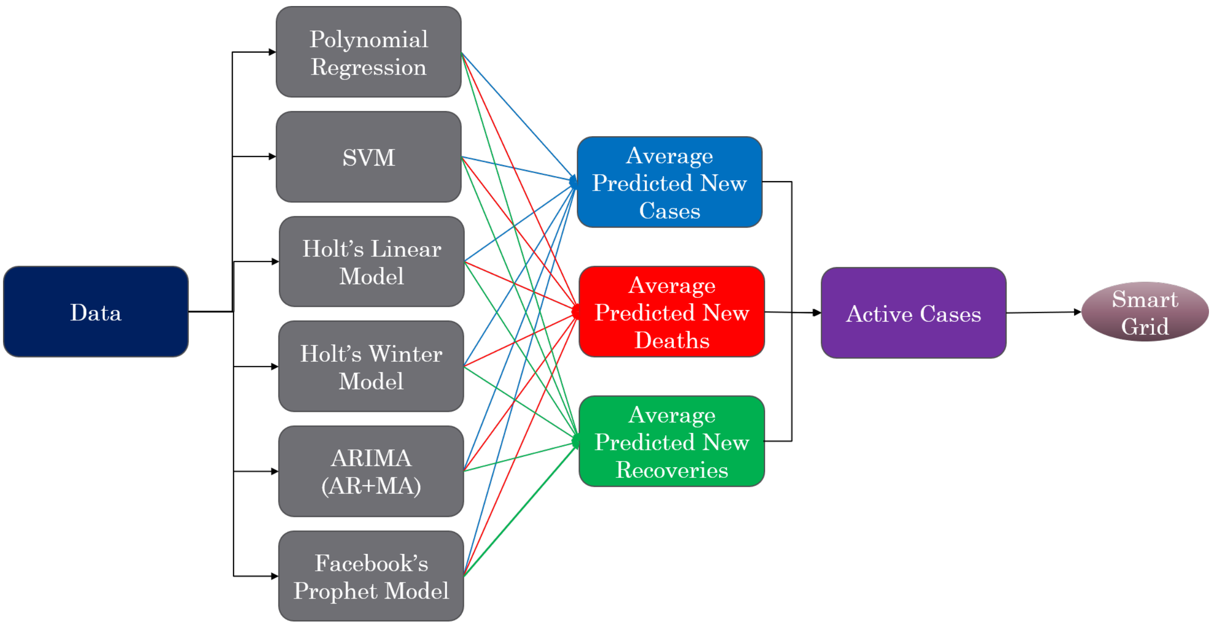


Figure 4.1: Flowchart of the Algorithms

This give us a clear idea of how our model is working with the algorithms implemented to the data available.
The data used in this research has been taken from GitHub which was collected by the very reliable Centre for Systems Science & Engineering (CSSE) Department at John Hopkins University. It is world renowned for the most reliable source & accurate data collection during COVID-19 pandemic.

This data is then separate data into 3 sections, one for every prediction that is is produced. So, this will give the number of affected cases, the number of deaths & the number of recovered cases. All 3 sections then are implemented to the model one by one, which then gives 6 individual predictions for each individual algorithms. This is then summed up & divided by 6 to get the Average Value of the specific prediction.

This way 3 average prediction values are produced, one for every section of data that we divided earlier for making the prediction. All these 3 predictions is used to find active cases which then is used to find out the necessary power demand.

4.2 Smart Grid System

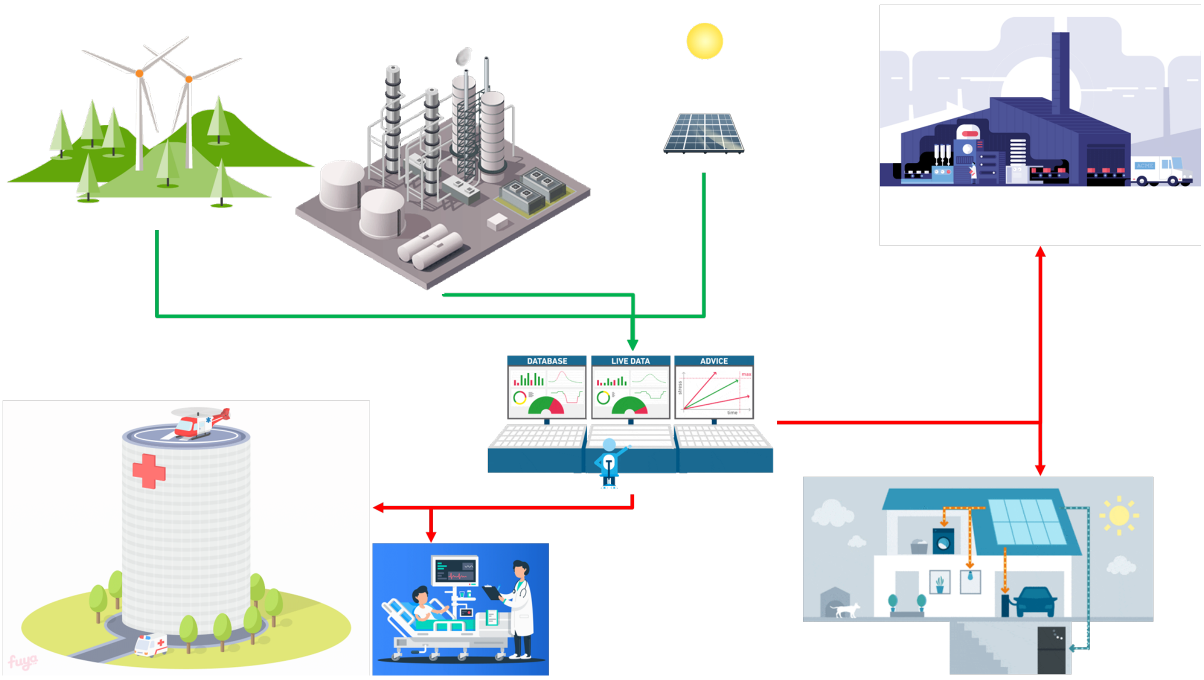


Figure 4.2: Smart Grid System

The current power distribution system is not capable enough to meet the excess demands that occur during a pandemic situation like COVID-19. During these time hospitals require extra power than usual due to excessive number of patients being admitted daily.

This research hopes to change that scenario by changing & upgrading our existing power distribution system.

1st of all, will incorporate more clean & renewable energy source like- solar panels, windmills, nuclear power plant etc. Which not only will give us more energy to work with but also will help us to save the mother earth for our future generation.

2nd of all, will incorporate a smart energy energy distribution hub. All the power generated will be under it's supervision. Also, the power demand forecast will be live broadcasted directly to this hub. Then using all these data & analysis the hub will distribute power based on priority. By default, hospitals are given the most importance as treating the affected patients is the main goal. After their demand the smart hub will provide power to residential & industrial areas.

3rd of all, will incorporate smart meters to all places where power is being distributed. This creates a two-way communication between the sender & receiver of power. Which is much more efficient than our existing power distribution system, as they are one-way power line. This helps us to know how much power they need & when they need it. As a result, a lot of power can be saved & used elsewhere.

Chapter 5

Evaluation

The implementation of different types of algorithms with the data collected from GitHub courtesy of John Hopkins University (CSSE Department), we get the full picture of COVID-19 situations in Bangladesh. Using these data, we also make a prediction & then find power demand which will be further discussed in this chapter.

5.1 Recovery & Mortality Rate

Recovery rate is the extent to which principal & accrued interest on defaulted debt can be recovered. It is expressed as a percentage of face value. Recoveries per 1000 individual per unit of time.

Mortality or death rate is a measure of the numbers of deaths (in general or due to a specific cause like COVID-19) in a particular population, scaled to the size of the population per unit of time. It is also expressed as a percentage face value. Deaths per 1000 individual per unit of time.

These values give us an idea of how a pandemic or epidemic is affecting countries. In case of COVID-19 we want a higher recovery rate & a mortality rate as low as possible. Because this means that for a specific amount of time, we are treating more people while losing less people per 1000. Which is a positive sign & really gives us courage while tackling a pandemic such as COVID-19.

As we can see from the figure given, Recovery rate of Bangladesh has been increasing quite drastically. Which is always good to see, as this is an evidence of people getting cautious & being smarter with dealing COVID-19; if affected. They are not panicking & following protocols to be in quarantine for 14 days and recovering from COVID.

And for the mortality rate, it slowly decreased from being at a scary high at the beginning & now has reached to a saturation point. This is also a reassuring sign to know that less people are dying everyday even though the number of affected people is still

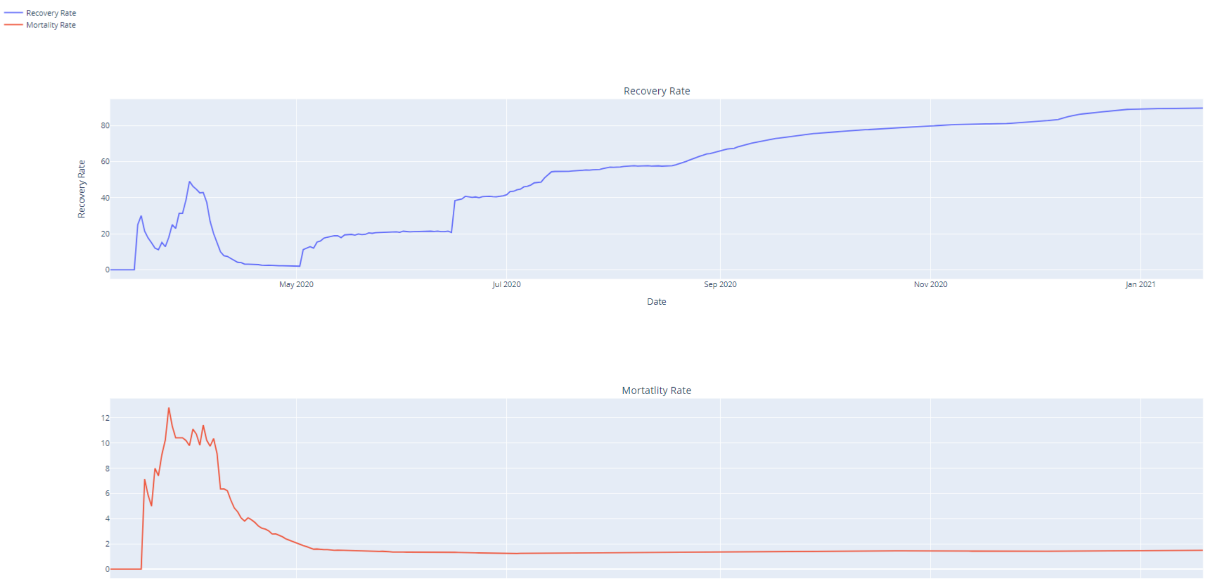


Figure 5.1: Recovery & Mortality Rate of Bangladesh for COVID-19

quite high.

5.2 Active & Closed Cases

Here we find two new terms-
i. Active Case- Total number of people being affected each day and number of people still recovering after being affected by COVID-19. This term gives us an idea of how many people are currently being affected by COVID-19 & are need of proper treatment.
Active Cases = No of total affected people

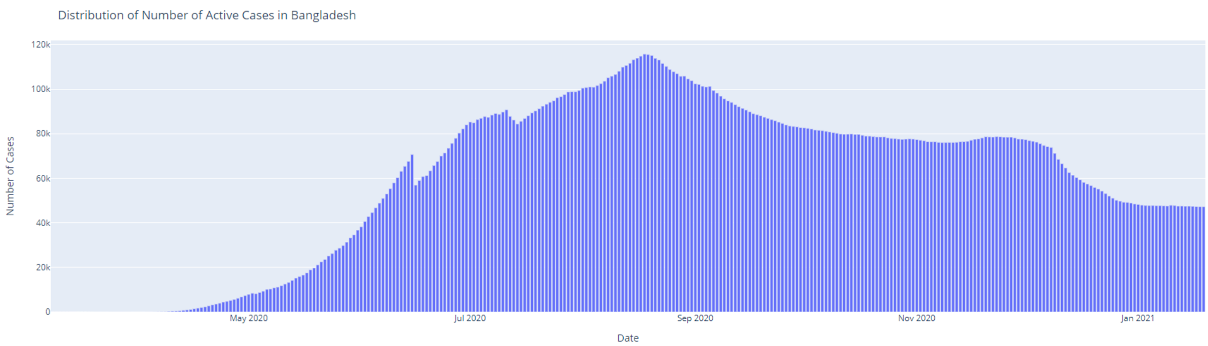


Figure 5.2: Active Cases of Bangladesh for COVID-19

ii. Closed Case- Total number of people that have died or are dying each day and the number of people that have recovered from COVID-19. This term gives us an idea of how many people are being treated & how many people currently out of the danger of spreading COVID.

Closed Cases = No of deaths + No of recovered

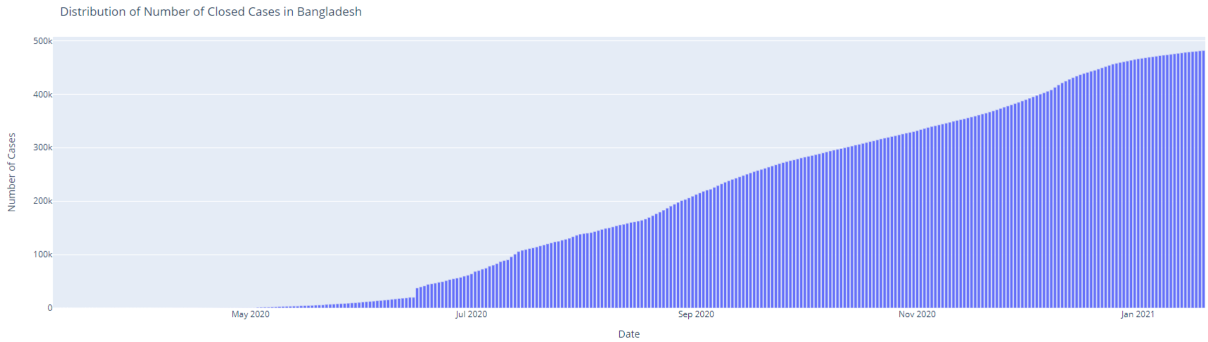
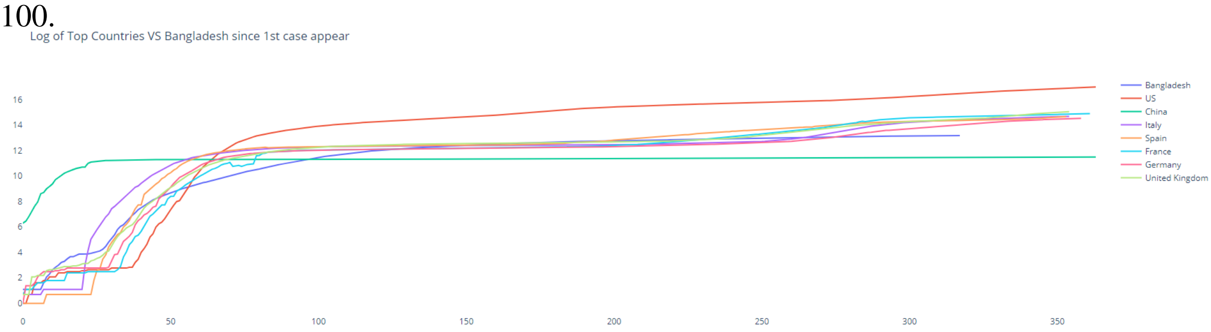


Figure 5.3: Closed Cases of Bangladesh for COVID-19

As we can see from the graphs, total number of active cases have gone up significantly in the 1st half. Which is very alarming to see & is a clear indication of COVID-19 spreading like crazy. This means that we have failed to abide by the Corona protocols & have put the people's lives at risk. But it has slowed a lot after September & is slowly on the decrease as many of us have already suffered and recovered from COVID-19. For Closed case, we necessarily do not want it to be as high as possible. Because this number being quite large does not always translate to better results. If we look closely, closed cases also involve number of people being dead. Which is not something we look for. But in general, having a higher value for closed case is ideal as this means we have a smaller number of people spreading the COVID germs. We can also interpret this smaller value as more people being treated & recovering from COVID-19.

5.3 Logarithmic comparison of Bangladesh with more developed countries

A logarithmic scale (or log scale) is a way of displaying numerical data over a very wide range of values in a compact way—typically the largest numbers in the data are hundreds or even thousands of times larger than the smallest numbers. Such a scale is nonlinear: the numbers 10 and 20, and 60 and 70, are not the same distance apart on a log scale. Rather, the numbers 10 and 100, and 60 and 600 are equally spaced. Thus, moving a unit of distance along the scale means the number has been multiplied by 10 (or some other fixed factor). Often exponential growth curves are displayed on a log scale, otherwise they would increase too quickly to fit within a small graph. Another way to think about it is that the number of digits of the data grows at a constant rate. For example, the numbers 10, 100, 1000, and 10000 are equally spaced on a log scale, because their numbers of digits is going up by 1 each time: 2, 3, 4, and 5 digits. In this way, adding two digits multiplies the quantity measured on the log scale by a factor of 100.



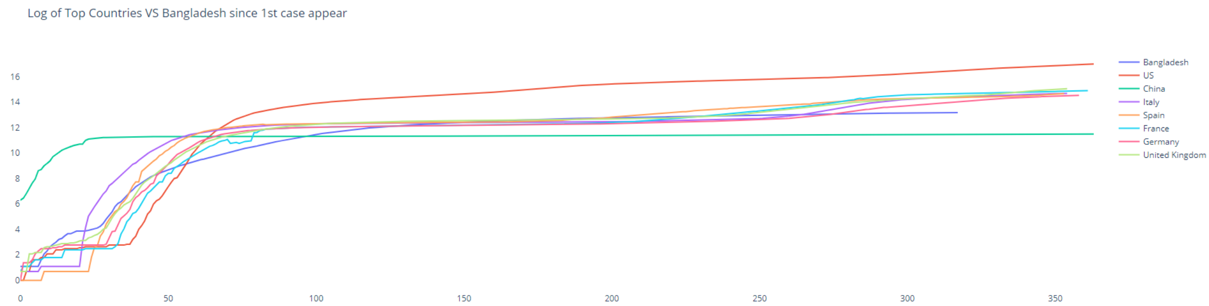


Figure 5.4: Logarithmic Scale of COVID-19 Patients in Bangladesh Compared To Developed Countries

This figure illustrates how the number of affected cases of COVID-19 has grown after the 1st infected case was discovered. This gives us a good idea of how each country handled the pandemic situation after discovery. And as we can see, all of the countries except China had fewer cases at the beginning. That is because China was the 1st country to be affected by COVID-19 as it was the birthplace. So, they did not have proper time & knowledge to prepare for such a global pandemic that the world has seen yet. But it is worth to note that now China has the lowest number of affected cases, which shows their tenacity & hard-working ethics to tackle such a massacre and still come out on top as victorious.

As for Bangladesh, we had 3 extra months to prepare for such global pandemic after seeing the horror scenes from China. And we did have a lower number of cases at the beginning like many other countries. But after the initial period the number shot up & sky-rocketed through the roof. We should have been able to keep the number much lower considering all the warnings & the extra 3 months we got to prepare for COVID-19.

5.4 Predicted Results

In order to predict the power requirement, we fist need to predict how much patients we need to treat to make an accurate assumption.
We make 3 separate predictions (Affected, Death, Recovered Cases) using all the algorithms to predict the actual number of COVID-19 patients.

5.4.1 Predicted Affected Cases

Date	Average Predictions of Affected Cases	Daily New Positive
20/01/2021	577780	0
21/01/2021	580189	2408
22/01/2021	582557	2369
23/01/2021	584884	2327
24/01/2021	587209	2325
25/01/2021	589661	2452
26/01/2021	592156	2494
27/01/2021	594663	2508
28/01/2021	597255	2592
29/01/2021	599693	2437
30/01/2021	602193	2500
31/01/2021	604776	2583
01/02/2021	607390	2614
02/02/2021	610012	2622
03/02/2021	612743	2731
04/02/2021	615458	2716
05/02/2021	618229	2770

Table 5.1: Average Prediction of Affected Cases for all 6 algorithms

5.4.2 Predicted Death Cases

Date	Average Predictions of Death Cases	Daily New Deaths
20/01/2021	8758	0
21/01/2021	8804	46
22/01/2021	8849	45
23/01/2021	8894	46
24/01/2021	8940	45
25/01/2021	8986	46
26/01/2021	9032	46
27/01/2021	9078	46
28/01/2021	9123	46
29/01/2021	9168	45
30/01/2021	9213	45
31/01/2021	9258	45
01/02/2021	9303	45
02/02/2021	9348	45
03/02/2021	9393	44
04/02/2021	9438	45
05/02/2021	9481	43

Table 5.2: Average Prediction of Death Cases for all 6 algorithms

5.4.3 Predicted Recovered Cases

Date	Average Predictions of Recovered Cases	Daily New Recoveries
20/01/2021	541813	0
21/01/2021	545475	3661
22/01/2021	549046	3572
23/01/2021	552628	3582
24/01/2021	556260	3632
25/01/2021	560062	3801
26/01/2021	563946	3884
27/01/2021	567951	405
28/01/2021	571848	38972
29/01/2021	575749	3901
30/01/2021	579658	3909
31/01/2021	583687	4030
01/02/2021	587773	4085
02/02/2021	591967	4194
03/02/2021	596201	4233
04/02/2021	600519	4319
05/02/2021	604727	4208

Table 5.3: Average Prediction of Recovered Cases for all 6 algorithms

As we know, we need 3 predictions to make a total estimation of how many patients are active at a certain time with COVID-19 virus that need proper treatment.
From all these predictions we also find the daily new cases (Daily New Active, Daily New Death, Daily New Recovered) for each one of the scenarios. And this is very easy as we only need to subtract the previous day case number from the case number of that specific day.
Daily New Cases (Active, Death, Recovered) = Total Predicted Cases Today – Total Predicted Cases for the Previous Day

5.5 Hospital Bed Required

In order to find out the total hospital bed we need for a specific day we implement a term called 'Active Cases' like we mentioned earlier. The only difference this time is that now the number of Active Cases is for predicted value & not actual values that have already been recorded.
So, the value of 'Predicted Active Case' will be the number of hospital bed requirement if we want to treat all of the patients currently suffering from the COVID-19 virus.

Predicted Active Cases = Average Prediction of Affected - Average Prediction of Deaths - Average Prediction of Recovered

Hospital Bed Required = Predicted Active Case

Date	Hospital Bed Required	Daily Power Demand (kWh)	Extra Bed Required
20/01/2021	27264	545287	0
21/01/2021	1341	26812	2896
22/01/2021	1302	26034	2935
23/01/2021	1289	25788	2948
24/01/2021	1396	27928	2841
25/01/2021	1321	26414	2916
26/01/2021	1428	28565	2809
27/01/2021	1448	28964	2789
28/01/2021	1525	30499	2712
29/01/2021	1307	26146	2930
30/01/2021	1549	30980	2688
31/01/2021	1451	29023	2786
01/02/2021	1651	33025	2586
02/02/2021	1503	30065	2734
03/02/2021	1492	29841	2745
04/02/2021	1564	31286	2673
05/02/2021	1637	32748	2600

Table 5.4: Hospital Bed Requirement & Daily Power Demand

As we can see, if we want to treat all of the people affected by COVID-19 or such a virus in the future; the amount of bed required to do so can easily be predicted by our model. And for every bed (ICU & non-ICU units) we need an average power of 20 kWh. This value may change depending upon the hardware needed in order to treat a patient properly as we may need more advanced & sophisticated equipment in the future for other pandemic scenarios. But at the moment using traditional equipment we can estimate the power needed for COVID-19 patients.
From IEDCR we find the total number of hospital beds (ICU & non-ICU units) available at Bangladesh is 4592. Just by subtracting this value from the 'Hospital Bed Required' we can find out 'Extra Bed Required' for us to completely fulfil our dreams of providing everyone proper treatment using hospital facilities.

5.6 RMSE

The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed.
The RMSE represents the square root of the second sample moment of the differences between predicted values and observed values or the quadratic mean of these differences. These deviations are called residuals when the calculations are performed over the data sample that was used for estimation and are called errors (or prediction errors) when computed out-of-sample. The RMSE serves to aggregate the magnitudes of the errors in predictions for various data points into a single measure of predictive power. RMSE is a measure of accuracy, to compare forecasting errors of different models for a particular dataset and not between datasets, as it is scale-dependent.

RMSE is always non-negative, and a value of 0 (almost never achieved in practice) would indicate a perfect fit to the data. In general, a lower RMSE is better than a higher one. However, comparisons across different types of data would be invalid because the measure is dependent on the scale of the numbers used.
RMSE is the square root of the average of squared errors. The effect of each error on RMSE is proportional to the size of the squared error; thus, larger errors have a disproportionately large effect on RMSE. Consequently, RMSE is sensitive to outliers.

$$RMSE = \sqrt{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2/N} \tag{5.1}$$

where,

$RMSE$ = Root Mean Square Error

i = Variable i

n = Number of non-missing data points

x_i = Actual observations time series

$\hat{x}_i$ = Estimated time series

Model Name	RMSE Value
3. Holt's Winter Model	421.911212
5. Moving Average (MA)	732.060941
2. Holt's Linear Model	746.973482
6. ARIMA	828.978337
4. Auto Regressive (AR)	1063.171462
7. Facebook's Prophet Model	2675.153792
0. Polynomial Regression	7569.329970
1. Support Vector Machine Regressor	307025.219836

Table 5.5: RMSE Value of Predicted Positive Cases

Model Name	RMSE Value
6. ARIMA	22.256990
3. Holt's Winter Model	44.503528
4. Auto Regressive (AR)	45.511317
5. Moving Average (MA)	45.647394
2. Holt's Linear Model	52.728858
0. Polynomial Regression	54.883397
1. Support Vector Machine Regressor	4397.915359

Table 5.6: RMSE Value of Predicted Death Cases

Model Name	RMSE Value
2. Holt's Linear Model	1499.813632
3. Holt's Winter Model	2867.274984
4. Auto Regressive (AR)	6067.04402
5. Moving Average (MA)	6067.04402
0. Polynomial Regression	38789.497778
6. ARIMA	242069.469826
1. Support Vector Machine Regressor	292470.351965

Table 5.7: RMSE Value of Predicted Recovered Cases

The lower the RMSE value of an algorithm, the better it is suited for our model.

As we can see, not all algorithms produce the best result possible. There are a lot of error due to the data.
One might wonder, why use algorithms with higher RMSE value which produces less than ideal values for our model. And the answer would be very simple, one algorithm might work better for our dataset but might not be suitable for other country's dataset. So, instead of finding a particular algorithm best suitable for each & every country we make a global model with various algorithms which are typically suited for our time series prediction model datasets.
This solves our problem of trying to find an individual algorithm for a specific country by pain stinking trial & error method. Thus, saving us hours trying to fit the best algorithm according to our dataset. We can rest assured & just easily change the dataset from Bangladesh to the country which we want to figure out the COVID-19 scenario.

Chapter 6

Discussion

This study has investigated about time series prediction models & epidemiological prediction models which can be used for an upcoming global pandemic like COVID-19. COVID-19 has exponentially & quickly spread to more than 200 countries worldwide with more than 36 million confirmed cases. [21] [22] [23] Regardless of what one's beliefs are, we believe that forecasts and their associated uncertainty can and should be an integral part of the decision-making process, especially in high-risk cases. [2] Forecasting the outcome of outbreaks as early and as accurately as possible is crucial for decision-making and policy implementations. [24] [25] As an alternative to epidemiologic transmission model, we used MAE to forecast the real-time trajectory of the transmission dynamics and generate the real-time forecasts of Covid-19. The MAE models allow inputting the interventions information and investigating the impact of interventions on the size of the virus outbreak and end time of the virus outbreak. [3] Two epidemiological models - a simple exponential model and an SIR model, are used respectively to forecast short- and long-term outcomes. These models assume all the seed cases to be symptomatic, which may underestimate the actual numbers due to an uncertain number of asymptomatic individuals. [4] Forecasting is exceptionally vital even to get the slightest result for multi variables consideration over public health factors, especially pandemic crises like COVID-19. In this case, forecast from single models is not enough for reliable results and prediction. Time series models provide a different and unique approach to time series forecasting. Basically, for the time series forecast, two approaches are widely used, i.e., exponential smoothing and Time Series Models like ARIMA and Richard's. While exponential smoothing models are based on a description of the trend and seasonality in the data, Time Series, like ARIMA models, aims to describe the autocorrelations in the data. [19] Six machine learning approaches named CUBIST, RF, RIDGE, SVR, and stacking-ensemble learning, as well as ARIMA statistical model, were employed in the task of forecasting one, three, and six-days-ahead the COVID-19 cumulative confirmed cases in ten states with a high daily incidence. [26] The autoregressive distributed lag (ARDL) model, reveals that a long-run relationship holds between the COVID-19 cases and energy consumption and that the COVID-19 cases have a positive effect energy consumption. [27] The use of dynamic time warping allows to identify similar, but time-shifted, time series. [28] The global pandemic has also made an impact on the overpopulated developing country Bangladesh. In Bangladesh, the Institute of Epidemiology, Disease Control and Research (IEDCR) has reported the first 3 cases of coronavirus on 8 March 2020. Due to the lacking of the real data related to the situation, it is a challenging task to estimate the parameters accurately. The best solution of the situation can only be a lockdown situation where everyone can be kept distant from infected individuals. [6] The still escalating COVID-19 pandemic also has a substantial impact on energy structure, requirements and related emissions. [29] [27] The demand of electricity has been reduced significantly due to the recent COVID-19 pandemic. This changed the lifestyle globally as people are mostly staying home and working from home if possible. Hence, there is a significant increase in residential load demand while there is a substantial decrease in commercial and industrial loads. [30]

Chapter 7

Conslusion & Future Work

In this chapter we discuss about how to solve the shortcomings & further future work that can be done to improve our model.

Improving the efficiency

From the result section we find our error, which is not ideal at any case. Because while forecasting such an important global issue we cannot afford to make any kind of mistakes, no matter how tiny it may be. We are predicting peoples lives here & deciding on how many people we can save so; it has to be accurate as possible.

As for improving our existing model, there are some things that can be implemented. We can try changing some or most of our algorithms implemented here & check how the new model reacts to the changes. We are confident of our existing model but if some other algorithm can make it better then we are all ears.

We have seen others use Bayesian Ridge Regression model which caught our eye but opt not to go for it this time around. It would be interesting to see if this can further improve our model.

Implementing the smart grid

For our model to change the current scenario of our power sector & meet the demands for an upcoming global pandemic we need better implementation of the smart grid system. Here we give an idea of how it can work but there is still a lot of questions to be answered & tests to be done. We need to fine tune & adjust how efficient it can be so that we do not need to make adjustments during a global pandemic like COVID-19.

REFERENCES

[1] L. Wynants, B. Van Calster, G. S. Collins, R. D. Riley, G. Heinze, E. Schuit, M. M. Bonten, D. L. Dahly, J. A. Damen, T. P. Debray *et al.*, "Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal," *bmj*, vol. 369, 2020.

[2] F. Petropoulos and S. Makridakis, "Forecasting the novel coronavirus covid-19," *PloS one*, vol. 15, no. 3, p. e0231236, 2020.

[3] Z. Hu, Q. Ge, S. Li, L. Jin, and M. Xiong, "Artificial intelligence forecasting of covid-19 in china," *arXiv preprint arXiv:2002.07112*, 2020.

[4] R. Ranjan, "Predictions for covid-19 outbreak in india using epidemiological models," *MedRxiv*, 2020.

[5] C. Massonnaud, J. Roux, and P. Crépey, "Covid-19: Forecasting short term hospital needs in france," *medrxiv*, 2020.

[6] M. S. Islam, J. I. Ira, K. Kabir, and M. Kamrujjaman, "Covid-19 epidemic compartments model and bangladesh," *Preprint, DOI*, vol. 10, 2020.

[7] R. R. Ibrahim and O. H. OLADIPO, "Forecasting the spread of covid-19 in nigeria using box-jenkins modeling procedure," *medRxiv*, 2020.

[8] E. Z. Martinez, D. C. Aragon, and A. A. Nunes, "Short-term forecasting of daily covid-19 cases in brazil by using the holt's model," *Revista da Sociedade Brasileira de Medicina Tropical*, vol. 53, 2020.

[9] N. Poonia and S. Azad, "Short-term forecasts of covid-19 spread across indian states until 1 may 2020," *arXiv preprint arXiv:2004.13538*, 2020.

[10] A. Argawu, "Modeling and forecasting of covid-19 new cases in the top 10 infected african countries using regression and time series models," *medRxiv*, 2020.

[11] S. Mahmud, "Bangladesh covid-19 daily cases time series analysis using facebook prophet model," *Available at SSRN 3660368*, 2020.

[12] K. Abdulmajeed, M. Adeleke, and L. Popoola, "Online forecasting of covid-19 cases in nigeria using limited data," *Data in brief*, vol. 30, p. 105683, 2020.

[13] S. Das, A. Dey, A. Pal, and N. Roy, "Applications of artificial intelligence in machine learning: review and prospect," *International Journal of Computer Applications*, vol. 115, no. 9, 2015.

[14] E. Grossi and M. Buscema, "Introduction to artificial neural networks," *European journal of gastroenterology & hepatology*, vol. 19, no. 12, pp. 1046–1054, 2007.

[15] T. Vijayapriya and D. P. Kothari, "Smart grid: an overview," *Smart Grid and Renewable Energy*, vol. 2, no. 4, pp. 305–311, 2011.

[16] G. Dileep, "A survey on smart grid technologies and applications," *Renewable Energy*, vol. 146, pp. 2589–2625, 2020.

[17] K. K. Zame, C. A. Brehm, A. T. Nitica, C. L. Richard, and G. D. Schweitzer III, "Smart grid and energy storage: Policy recommendations," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1646–1654, 2018.

[18] A. Ghasempour, "Internet of things in smart grid: Architecture, applications, services, key technologies, and challenges," *Inventions*, vol. 4, no. 1, p. 22, 2019.

[19] P. Kumar, R. K. Singh, C. Nanda, H. Kalita, S. Patairiya, Y. D. Sharma, M. Rani, and A. S. Bhagavathula, "Forecasting covid-19 impact in india using pandemic waves nonlinear growth models," *MedRxiv*, 2020.

[20] A. S. Ahmad, M. Y. Hassan, M. P. Abdullah, H. A. Rahman, F. Hussin, H. Abdullah, and R. Saidur, "A review on applications of ann and svm for building electrical energy consumption forecasting," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 102–109, 2014.

[21] A. Zeroual, F. Harrou, A. Dairi, and Y. Sun, "Deep learning methods for forecasting covid-19 time-series data: A comparative study," *Chaos, Solitons & Fractals*, vol. 140, p. 110121, 2020.

[22] I. Rahimi, F. Chen, and A. H. Gandomi, "A review on covid-19 forecasting models," *Neural Computing and Applications*, pp. 1–11, 2021.

[23] R. Gupta and S. K. Pal, "Trend analysis and forecasting of covid-19 outbreak in india," *MedRxiv*, 2020.

[24] F. Petropoulos, S. Makridakis, and N. Stylianou, "Covid-19: Forecasting confirmed cases and deaths with a simple time series model," *International Journal of Forecasting*, 2020.

[25] E. Gecili, A. Ziady, and R. D. Szczesniak, "Forecasting covid-19 confirmed cases, deaths and recoveries: Revisiting established time series modeling through novel applications for the usa and italy," *PloS one*, vol. 16, no. 1, p. e0244173, 2021.

[26] M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, and L. dos Santos Coelho, "Short-term forecasting covid-19 cumulative confirmed cases: Perspectives for brazil," *Chaos, Solitons & Fractals*, vol. 135, p. 109853, 2020.

[27] K. Aruga, M. Islam, A. Jannat *et al.*, "Effects of covid-19 on indian energy consumption," *Sustainability*, vol. 12, no. 14, p. 5616, 2020.

[28] J. Stübinger and L. Schneider, "Epidemiology of coronavirus covid-19: Forecasting the future incidence in different countries," in *Healthcare*, vol. 8, no. 2. Multidisciplinary Digital Publishing Institute, 2020, p. 99.

[29] J. J. Klemeš, Y. Van Fan, and P. Jiang, "The energy and environmental footprints of covid-19 fighting measures–ppe, disinfection, supply chains," *Energy*, vol. 211, p. 118701, 2020.

[30] R. M. Elavarasan, G. Shafiullah, K. Raju, V. Mudgal, M. T. Arif, T. Jamal, S. Subramanian, V. S. Balaguru, K. Reddy, and U. Subramaniam, "Covid-19: Impact analysis and recommendations for power sector operation," *Applied energy*, vol. 279, p. 115739, 2020.