Investigating the Effect of Government Policies on the Spread out of Covid-19

by

Maryam Yalsavar

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computational Mathematics

Waterloo, Ontario, Canada, 2023

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The COVID-19 outbreak has been identified as one of the most severe respiratory virus outbreaks since the 1918 H1N1 influenza pandemic. Its impact on people's lives and governments has been significant, affecting areas such as health and the economy. Governments have responded to the pandemic by implementing policies to mitigate its destructive effects on the economy and people's lives and to stop its spread. These policies fall under four main categories: containment and closure policies, economic policies, health system policies, and vaccination policies.

Many of these policies were implemented as emergency measures without a thorough study of their impact and effectiveness. Therefore, investigating the impact of government policies during the pandemic is crucial to assist policymakers in addressing future pandemics or variations of COVID-19.

To conduct this investigation, we utilized counterfactual reasoning and counterfactual generation techniques from causality. We also measured the level of dependence between each policy and the spread of COVID-19 using Hilbert-Schmidt Independence Criterion (HSIC) and mutual information. Our findings indicate that vaccination policies had the most positive impact on controlling the disease. Additionally, school closings, restrictions on gatherings, and canceling public events were found to be quite effective. However, in some cases, our methods produced counterintuitive results, suggesting a decrease in the level of some policies, such as PCR testing, for controlling the disease.

It's important to note that different methods used in this investigation may produce different and sometimes contradicting results. We discussed the limitations of the techniques used, which may have contributed to the contradictory findings.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Ali Ghodsi for his continuous support of my Master's research, for his patience, motivation, enthusiasm, and immense knowledge.

I would like to acknowledge Mr. Mazi Hosseini the manager of Arshon Technology Inc. for his generous supports and helps.

I would like to thank Prof. Hans De Sterck, Prof. Ryan Browne, Erin Kelly, Alicia Hanbidge and the rest of the Computational Mathematics faculty, staff, and students for all supports and to create a great learning environment and pleasant community to pursue my Master's degree.

Dedication

Behind all successful individuals and stars, there were people who believed them and supported them when they were losing their hopes and forgetting themselves. This is dedicated to the ones who sacrificed themselves for those who shed light on our unknowns.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

During the COVID-19 pandemic, governments implemented various policies to control the spread of the disease and mitigate its negative impacts on people's lives. These policies can be categorized into four main categories: containment and closure policies, economic policies, health system policies, and vaccination policies. However, many of these policies were implemented as emergency measures without a thorough study of their impact and effectiveness. Therefore, investigating the impact of government policies during the pandemic is crucial to assist policymakers in addressing future pandemics or variations of COVID-19.

Numerous studies have attempted to investigate the underlying effect of policies on COVID-19. Some studies used statistical methods and approaches for this purpose. For instance, in [8], the authors considered time, log(time), and the lagged value of different policies as features and found their coefficient in their Poisson regression model. Negative coefficients indicated that the feature had a positive effect, and vice versa. The authors found that most of the significant policies had positive coefficients, suggesting that daily confirmed cases and policies may be correlated in more complex ways or that other factors not considered in the study may affect the spread of COVID-19. The study also found that contact tracing and the health index had consistent negative coefficients across all analyses.

Other studies used causality to investigate the effect of policies. Some studies tried to find the causal graph to show which policies were the cause of the number of daily confirmed cases without indicating the positivity or negativity of the effect. For example, in [13], authors modified the SyPI algorithm [14] to generate the causal graph by considering confounders and generated the causal graph of the effect of different types of closures and the effect of different cities on each other in Germany. The study found that different cities had a different set of policies as the causes, and the set of causes was not the same for all cities.

Other studies assumed a causal graph between policies and the number of daily confirmed cases and used interventions to find the effects. For instance,in [19], they assumed a causal graph, derived the equations behind it, and found the coefficients of the policies on those equations to show the effect of those policies.

Additionally, a research introduced Simpson's paradox in COVID-19 case fatality rates and argued that a country is like a confounder that affects both age and fatality rate [20]. Ignoring the effect of this confounder when investigating the effect of policies will hide the true effect of age on the fatality rate.

This study aims to investigate the effect of policies on the number of daily confirmed cases during the COVID-19 pandemic. We assumed an underlying causal graph and employed deep learning/machine learning models for modeling the dynamic of the problem to apply counterfactual reasoning. We aimed to answer three questions: (1) What would have been the impact on the daily confirmed cases if the policies were not implemented or implemented with lower levels of strictness? (2) What is the optimal level of each policy to ensure that the number of daily confirmed cases remains below 500 during the peak days? (3) What is the level of dependence between policies and the number of daily confirmed cases? The report is organized as follows: Chapter 1 introduces the problem statement, Chapter 2 presents background information, Chapter 3 describes the proposed methods, Chapter 4 presents the experimental results, Chapter 5 discusses the limitations of the employed methods, and Chapter 6 concludes the paper.

Chapter 2

Background and Problem Definition

The coronavirus disease 2019 (COVID-19) has rapidly spread globally and was declared a pandemic by the World Health Organization on March 11, 2020 (World Health Organization, n.d.) [6]. The COVID-19 pandemic has had a significant impact on people's lives, world economies, and public health, and is considered the most severe respiratory virus outbreak since the 1918 H1N1 influenza pandemic (World Health Organization, n.d.)[7]. Governments worldwide have implemented various policies to mitigate or suppress the disease, which can be classified into four categories: C for containment and closure policies, E for economic policies, H for health system policies, and V for vaccination policies. However, the effectiveness of each policy in mitigating the spread of COVID-19 may differ. Therefore, it is crucial to identify which policy or policies are most effective in controlling the disease. Determining the effectiveness of each policy can assist policymakers in addressing future pandemics or variations of COVID-19. To investigate the impact of each policy on COVID-19 transmission, we employed counterfactual reasoning and counterfactual generation techniques from causality, while utilizing Hilbert-Schmidt Norm and mutual information to measure the level of dependence between each policy and the spread of COVID-19. Before introducing our approaches, it would be great to introduce the mentioned methods from the available literature.

2.1 Causality

Causality can be illustrated through an example of a student, Matth, who is taking a math course. The more hours he spends practicing math, the higher his score is likely to be. This indicates a relationship between these two events, where a change in the number of practicing hours causes a change in the score. This is known as causality or causation, which refers to the relationship between a cause and its effect. If we have two variables X, and Y then we say that X is a cause of Y if Y relies on X for its value[16]. We need causality and causation in machine learning due the obstacles that today's machine learning models are confronting [15]. The first obstacle is about adaptability or robustness, machine learning researchers has noticed that current systems lack the ability to recognize the new circumstances that they have not been trained for before, also machine learning models use historical data and as time passes some of the features may change so they require to be trained again. The second problem is about explain ability, machine learning models are mostly act as black boxes. So, they are unable to bring some reasons for their predictions and to get some recommendation. Imagine a person who has been applied for getting loan from a bank but his application was rejected by the system of that bank. The system must be able to provide some explanations on its decision and also provide some possible recommendations for that person to be able to get a loan. The third obstacle is about ignoring the cause-effect relations[12]. Machine learning models require to be able to consider these causal relations to be able to arrive at a human intelligence level while they are mostly operating based on statistics.

2.2 Counterfactual Reasoning (CR)

Counterfactuals are a fundamental concept in causality that aim to address questions of the form: "What would have happened if I had acted differently?" which the "if" portion is unrealistic or untrue. Interventions are not capable of answering such questions because counterfactuals consider events occurring under different conditions or in different worlds, which is not possible in interventions. To calculate counterfactuals, a structural causal model (SCM) is required, which illustrates how the variables in a given context interact and are related. An SCM comprises two types of variables: exogenous variables, which are unknown and act as external factors affecting other variables, denoted as U, and endogenous variables, which are descended from at least one of the exogenous variables[16]. In the case of our problem, we assume that policies have a direct effect on the number of daily confirmed cases and have constructed an SCM using deep learning models, as shown below:

$$y = f(x) + e \tag{2.1}$$

Where e is the exogenous variable, x is the set of features (policies) and f is the underlying function that we have found it by deep learning models and y is the number of daily confirmed cases. After discovering the underlying Structural Causal Model (SCM) of a problem, the following three steps must be taken to calculate counterfactuals[16]:

1. Abduction: Using the available observations for each instance to calculate its corresponding exogenous variables
2. Action: modify the underlying model (M) by replacing the intervened features ($x^{'}$) for arriving at a new model ($M_{x'}$)
3. Prediction: calculate the consequence of the counterfactual by using the modified model $M_{x'}$ and the corresponding exogenous variables that have been calculated in the Abduction step

These procedures can be demonstrated via the following example: consider a scenario where the number of daily confirmed cases of COVID19 depends on the level of strictness of three policies: health(h), vaccine(v) and closures(c). Let us assume that for $h = 30$, $v = 0$, and $c = 10$, $f(h = 30, v = 0, c = 10) = 19997$, and the recorded number of daily confirmed cases was around 20000. Based on the Abduction step, we find the exogenous variable, $e = 20000 - 19997 = 3$. Next, we intervene in the features such that $v = 0$, $h = 30$, $c = 100$, and find the modified model, $f_{v=0,h=30,c=100} = 10000$ based on the action step. Finally, the resulting counterfactual will be $y_{counterfactual} = 10000 + 3 = 10003$ which implies that the closure policy had a positive effect on controlling COVID19.

2.3 Counterfactual Generation (CG)

Machine learning models have become increasingly popular for solving various tasks such as classification, regression, and recommendation systems. However, one of the limitations of these models is their inability to provide recommendations to users on how to improve their state and achieve desired goals [16]. Counterfactual generation aims to address this limitation by generating recommendations that answer the question of how to change the current state of different features for an instance to transfer its state from A to B. Many works in the literature have addressed counterfactual generation to arrive at possible recommendations.

One current work [9], introduces a multi-objective optimization problem to cover different properties of the generated counterfactuals. According to their method, the generated counterfactual ($x^{'}$) should satisfy the following properties for its corresponding observation (x^{*}):

1. Its resulted output(state) is close to the desired output (state)
2. It is close to x^* in the input space (X)
3. It is different from x^* in a few features
4. It is a possible data point based on the distribution of input data (P_X)

To satisfy each of these properties, they define the objective function as follows:

$$\min_x O(x) = \min_x(O_1(\hat{f}(x), Y^{'}), O_2(x, x^*), O_3(x, x^*), O_4(x, X^{obs})) \tag{2.2}$$

In the context of counterfactual generation, four properties have been identified as important metrics to evaluate the quality of generated counterfactual instances. These properties are quantified by four separate objective functions, denoted as O_i , $i = 1, \ldots, 4$. A detailed description of each objective function is available in the literature. Here, we focus on the fourth objective function, which we use in our counterfactual reasoning method to measure the distance of generated counterfactuals from the distribution of the dataset

The fourth objective measures weighted average Gower distance[11] between x (the input) and the k nearest observed data points, $x[1], ..., x[k] \in X$. This empirical approximation helps determine the likelihood of x originating from the distribution of the primary data (X) as follows:

$$O_4(x, X^{obs}) = \sum_{i=1}^{k} w^{[i]} \frac{1}{p} \sum_{j=1}^{p} \delta_G(x_j, x_j^{[i]}) \in [0, 1] \quad where \sum_{i=1}^{k} w^{[i]} = 1 \tag{2.3}$$

And δ_G defines as follows:

$$\delta_G(x_j, x_j^*) = \begin{cases} \frac{1}{\hat{R}_j}|x_j - x_j^*| & \text{if } x \text{ is a numeriacl feature} \\ I_{x_j \neq x_j^*} & \text{if } x \text{ is a categorical feature} \end{cases} \tag{2.4}$$

is extracted from the dataset based on the value range of the j-th feature.

2.4 Measuring Dependency

Measuring the dependency between each feature, i.e., policies, and the target variable, i.e., the number of daily confirmed cases, is essential for revealing their effect and relation. To accomplish this, we employed the following two methods:

Hilbert-Schmidt Independence Criteria (HSIC)

This method is a kernel-based approach used to measure the dependency between variables. The underlying idea is that while measuring correlation in a linear space is straightforward, the unknown non-linearity of the data makes it challenging to measure the dependence factor. Therefore, based on a theorem, if we implicitly transform the data to a high-dimensional space using a kernel, such as the radial basis function (RBF), then the variables in the infinite-dimensional Reproducing Kernel Hilbert Spaces (RKHS) are independent if and only if they are uncorrelated. As a result, it is sufficient to find the correlation between the kernel means using kernel trick. For a set of independent observations that have been sampled from a distribution P_{xy} as $O := (x_1, y_1), \ldots, (x_m, y_m) \subseteq \times$, where m is the number of samples, HSIC can be calculated as follows[5]:

$$HSIC(O, F, g) := (m-1)^{-2} trKHLH \tag{2.5}$$

Where $H, K, L \in R^{m \times m}$, $K_{ij} := k(x_i, x_j)$, $L_{ij} := l(y_i, y_j)$ and $H_{ij} := \delta_{ij} - m^{-1}$

Mutual Information (MI)

Mutual information is a measure based on the entropy of variables that quantifies the amount of information that one variable provides about another variable. MI is a non-parametric measure of the dependence between two variables and ranges from 0 (indicating independence) to positive infinity, where a value of 0 indicates independence between two random variables, and higher values indicate stronger dependence between them. Mathematically, the mutual information between two random variables X and Y can be computed as follows[4]:

$$I(X; Y) = H(X) - H(X|Y) \tag{2.6}$$

Where I is the mutual information between X and Y, $H(X)$ is the entropy of X and $H(X|Y)$ is the conditional entropy of X given Y.

2.5 Deep Learning/ Machine Learning Models

For finding the underlying model of the COVID19 data, we have used Support Vector Regression(SVR)[3] and a Gated Recurrent Unit (GRU)[10] based network, that their details is mentioned bellow.

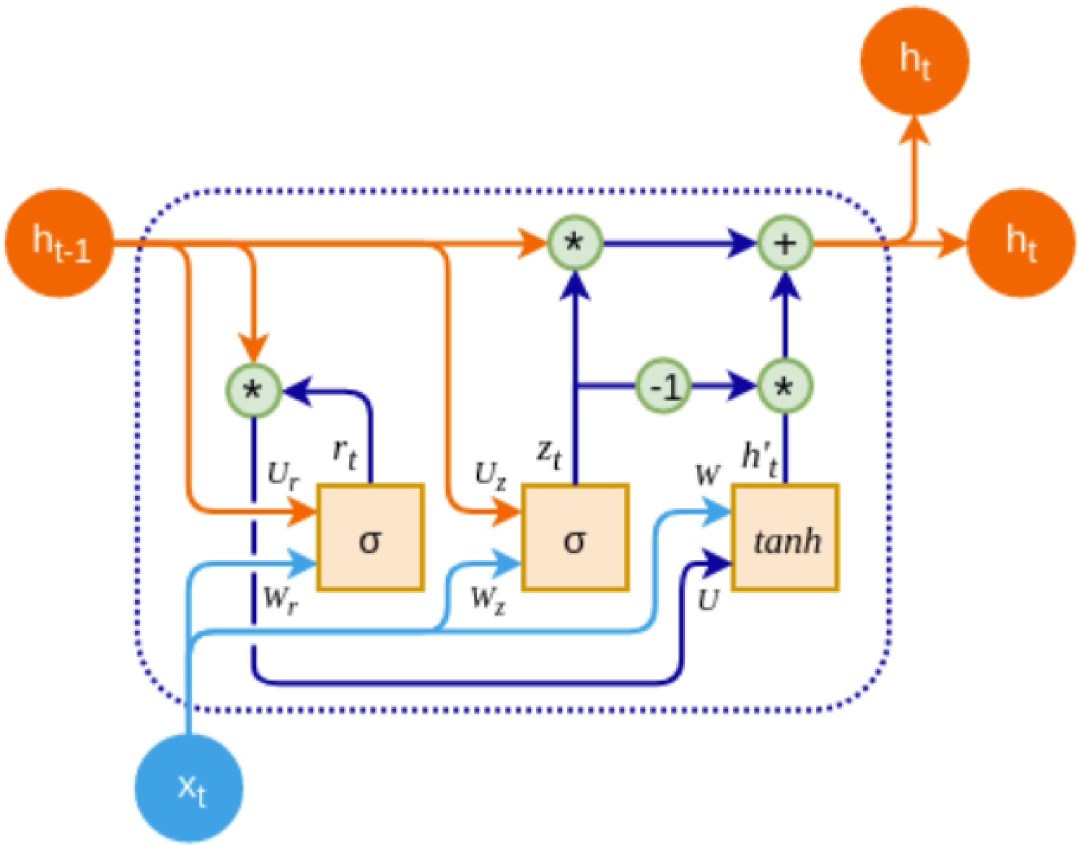

Figure 2.1: Gated recurrent unit (GRU) architecture[2].

Gated Recurrent Unit (GRU)

Recurrent Neural Networks (RNNs) are prone to the problem of gradient vanishing and bottleneck, which results in the loss of important information over long input sequences. In order to solve this issue, the Gated Recurrent Unit (GRU) was introduced[?]. GRU utilizes update and forget gates in its structure to address the problem of vanishing gradients and information loss in RNNs. The functional representation of a GRU unit can be seen in Figure 2.1 and is described below.

The update gate (z_t) determines which information needs to be kept and which needs to be discarded using the following formula:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{2.7}$$

The reset gate (r_t) determines how much of the previous state should be passed using the following formula:

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{2.8}$$

The memory content uses the reset gate to store relevant information from the past:

$$h_t^{'} = \tanh(W x_t + U(r_t * h_{t-1}) \tag{2.9}$$

Finally, the unit calculates the h_t which contains information about the current state, to be passed on to the next units using the following formula:

$$h_t = z_t * h_{t-1} \oplus (1 - z_t) * h_t^{'} \tag{2.10}$$

Support Vector Regression (SVR)

This is a method used for regression tasks where the labels are continuous. The goal is to find the best line by solving one of two optimization problems. The first optimization problem is[18]:

The first optimization problem is:

$$min \frac{1}{2} \|w\|^2 \quad s.t. |y_i - w_i x_i| \leq \varepsilon \tag{2.11}$$

The second optimization problem is:

$$min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} |\xi_i| \quad s.t. |y_i - w_i x_i| \leq \varepsilon + |\xi_i| \tag{2.12}$$

Here, ε represents the margin of error, y_i is the label, x_i is the input and w_i are coefficients.

The deviation from the margin ε is represented by ξ_i, which includes any error outside of the margin for some points from the line.

Chapter 3

Methodology

In this report, we aim to address the problem of determining the impact of policies on the transmission of Covid-19 by answering the following three research questions:

1. What would have been the impact on the daily confirmed cases if the policies were not implemented or implemented with lower levels of strictness?
2. What is the optimal level of each policy to ensure that the number of daily confirmed cases remains below 500 during the peak days?
3. What is the level of dependence between policies and the number of daily confirmed cases?

To answer the first two questions, we employ counterfactual reasoning/generation, while we estimate mutual information between each policy and the number of daily confirmed cases for the third research question. Further details of the employed methods can be found in subsequent sections.

3.1 Counterfactual Reasoning

To address the initial inquiry, we have employed counterfactual reasoning, a concept in causality, as a methodological approach. In order to implement this approach, it was necessary to develop a model of the problem. Therefore, we began by utilizing deep/machine learning models to model the effect of policies on the number of daily confirmed cases. Two

Sequence length	Train MSE		Test MSE		#train data	# test data
	GRU_model1	GRU_model2	GRU_model1	GRU_model2		
5	0.32230179930867514	0.4346392810129612	1.2528273836513348	1.2129056446786128	27158	6892
7	0.2987417784043683	0.3285427176648504	1.1543941183572677	1.095605739265928	27070	6859
10	0.210973340648669	0.29731745162944606	0.866611977065608	0.9799715153751649	26934	6817
15	0.15339124191113784	0.17105107342098852	0.6234000675083087	0.6291053129615952	26686	6770
30	0.01111500333849858	0.012078098885689207	0.01610591501799095	0.01885932536314295	25990	6581

Table 3.1: Mean square error and number of data for both train and test phases of GRU based models.

models, namely a GRU network and Support Vector Regression (SVR), were employed for this purpose.

In case of GRU network, two different GRU networks were trained using 5 distinct sequence lengths each. The mean of the results from all 10 models was reported as the final conclusion in order to mitigate the effects of any potential model bias on the results. A window with a length of 31 was utilized to represent the time required to observe the effect of a policy on the number of daily confirmed cases. Subsequently, 5, 7, 10, 15, and 30 were evaluated as possible sequence lengths, and for each input sequence, the corresponding y value at the end of the window was considered as the label. The process of generating the data can be observed in Figure 3.1. In the case of SVR, the model is not designed for modeling sequential data. Therefore, we input a set of features (policies) at each time and select the labels by considering a window of 30, 35, 40, 45, and 50. This window allows us to observe the effect of policies over time. To report our results, we calculate the average of the outcomes across these different window sizes to reduce model bias. To train these models, we used 80% of the data for training and validation purposes (70% for training, and 10% for validation). The remaining 20% of the data was used as the test set. To minimize the impact of seasonal patterns, we separated the data as shown in Figure 3.2. Finally mean squared error(MSE) along with the prediction curves on both train and test data have been reported in Tables 3.1 and 3.2 and Figures 3.3 and 3.4 as a metric for evaluating the generalizability of the models.

Following the development of the underlying data model, counterfactual reasoning can be applied through three distinct steps[16]: abduction, action, and prediction. In the first step, abduction, the unknown exogenous variable (e) is identified for each data point within the dataset, with Equation 1 representing the mathematical model of our problem. In this equation, y represents the daily confirmed cases and the label of each data point, $f(x)$ denotes the model prediction (utilizing either GRU networks or SVR) for a given input, x, and x represents the set of features, including policies and y_{ave}.

Sequence length	Train MSE	Test MSE	# train data	# test data
30	0.4419261780802710	0.7588354142474789	25990	6581
35	0.49529740520637894	0.8600352649315898	25742	6534
40	0.5390001543702576	0.945128562432032	25518	6463
45	0.5716183213149735	0.9831645856910654	25270	6416
50	0.593433933264728	1.0317995225237289	25046	6345

Table 3.2: Mean square erreor and number of data for both train and test phases of SVR models.

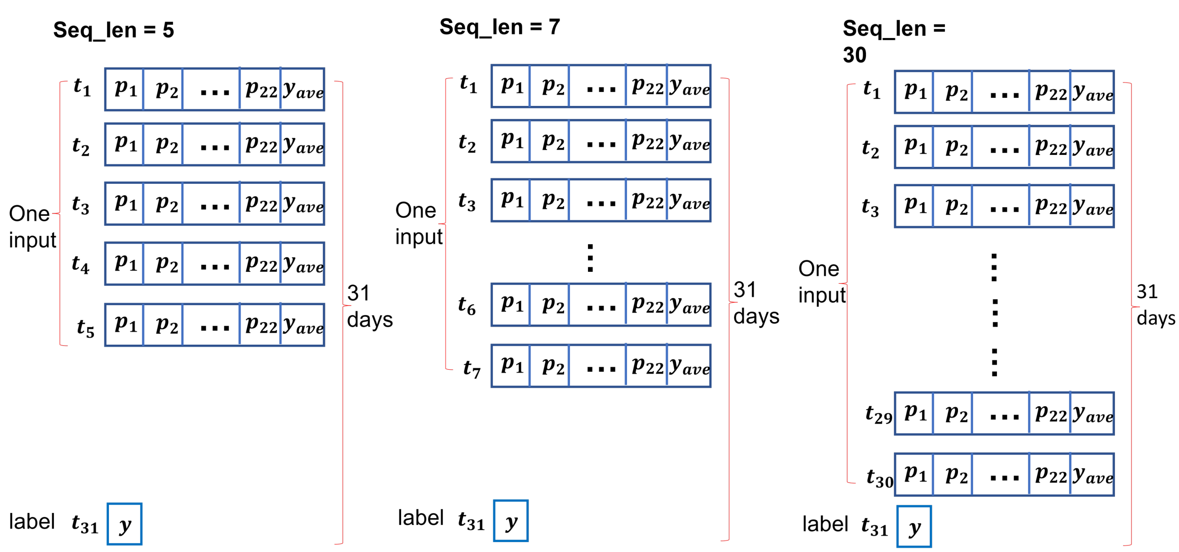

Figure 3.1: The process of generating sequential data (i.e., 5, 7, 10, 15, and 30) while a window length of 31 has been utilized to capture the time required to observe the effect of a policy on the number of daily confirmed cases.

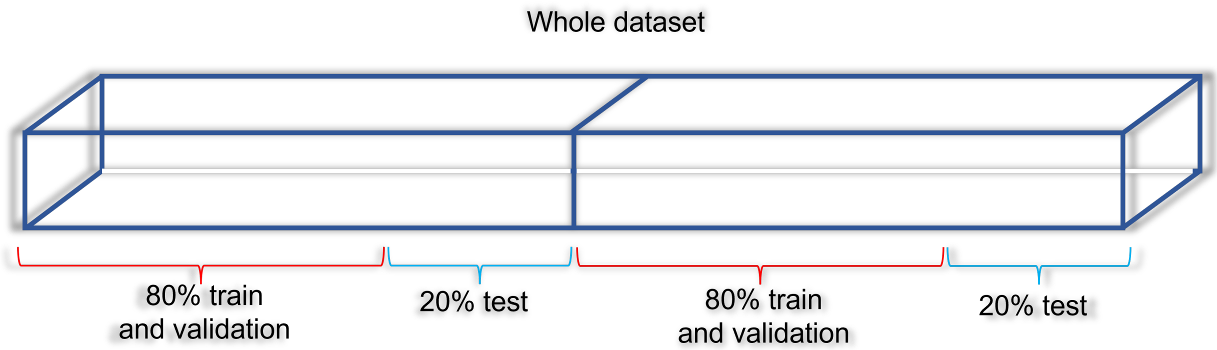

Figure 3.2: Data separation into train/validation/test for reducing seasonality patterns.

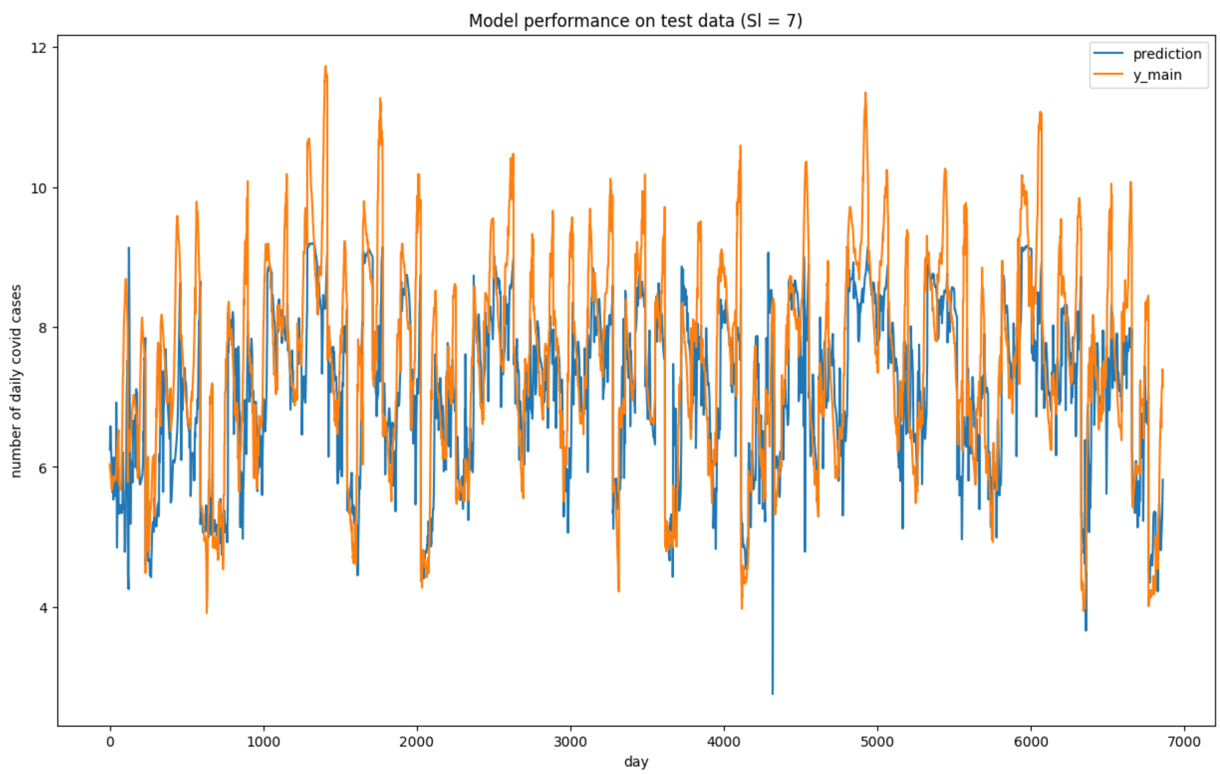

Figure 3.3: A sample on GRU model predictions on test data when sequence length is 7.

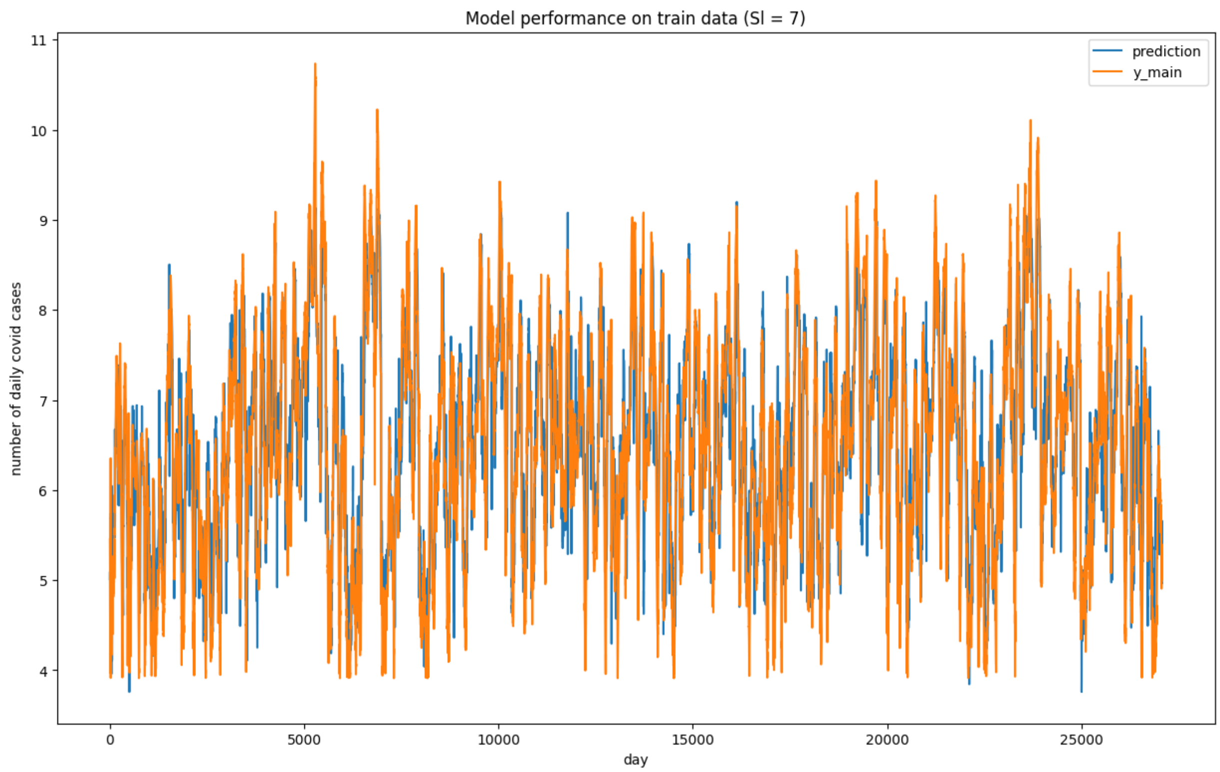

Figure 3.4: A sample on GRU model predictions on train data when sequence length is 7.

$$y = f(x) + e \tag{3.1}$$

In the second step, action, we replace the exogenous variable (e) in Equation 3.1 with the corresponding value for each data point. Next, we modify the input of $f(x)$ in the following manner: As our objective is to determine the effect of each policy, we set the value of a given policy to zero (i.e. $do(x = x^{'})$) and subsequently calculate the new f value ($f_{(}x^{'})$). Essentially, by setting the value of x to $x^{'}$, we are posing a counterfactual question to the model (GRU or SVR). However, utilizing an input that is from a distribution different from that of the training or test dataset can negatively impact the performance of neural network models such as GRU. To ensure that our counterfactual questions and samples are comparable to those in the training and test sets, we utilized the weighted average Gower distance[11] between x (the input) and the k nearest observed data points, $x[1], ..., x[k] \in X$ (Susanne Dandl, 2020). This empirical approximation helps determine the likelihood of x originating from the distribution of the primary data (X). Given that the input of GRU is a sequential data with a set of points, we determine the midpoint in each input and utilize the aforementioned measure to ensure that the counterfactual points are proximate to the dataset distribution.

Finally, in the prediction step, we calculate the consequences of the counterfactuals for each data point by utilizing its corresponding e and $f_{(}x^{'})$. By finding the difference between the number of daily confirmed cases before and after implementing the counterfactuals (eq.3.2), we can determine the impact of policies. As we set the value of policies to zero in our counterfactual questions, a positive difference indicates that the policy was effective in controlling the disease, while a negative difference suggests the opposite. Furthermore, the magnitude of this difference can be utilized to compare the effectiveness levels of different policies. After determining the differences for all data points using all models, we computed the mean of those results in the final report to facilitate interpretations and mitigate the impact of model bias. The mean of differences for each policy and for each model (GRU or SVR) is presented in Table 4.2.

$$difference = y_{conterfactual} - y \tag{3.2}$$

3.2 Counterfactual Generation

During the COVID-19 pandemic, governments have implemented various policies with varying levels of strictness in response to conditions, interpretations, and forecasts. Now we would like to go back to that environment in the past and ask this question: "What is the optimal level of each policy to ensure that the number of daily confirmed cases remains below 500 during the peak days?" It is not possible to change the level of policies in the past to see its effect on the number of daily confirmed cases because that state is past in time and we do not have access to that, but counterfactual generation gives us the ability to answer this question by using the available observational data. So, in this study, we employ the approach outlined in Susanne Dandl's 2020 paper to generate counterfactuals that reduce the number of daily confirmed cases during peak days to below 500. For this approach, we do not normalize policies or the number of daily confirmed cases, with policy levels ranging from 0 to 100 and the number of daily confirmed cases ranging from 0 to large positive numbers. By employing this method, we can calculate the change in the level of policies. It is expected to see a recommendation for increasing in the level of effective policies and a decrease in the level of those that were not effective, and the range of changes can state which policies were more important than others.

3.3 Calculating the Dependency Level Between Policies and Number of Daily Confirmed Cases

In this study, two methods, namely Mutual Information (MI) and Hilbert-Schmidt Independence Criterion (HSIC), were used to measure the dependency between policies and daily number of confirmed cases. While these methods cannot explicitly determine whether the policies were effective or not, they can provide an insight into the degree of effect that a policy may have had on the output variable. The obtained results for these methods are presented in Table 4.2.

Chapter 4

Experiments and Results

The dataset of policies, both numeric and ordinal, is sourced from the Oxford COVID-19 Government Response Tracker (OxCGRT)[1] developed by the Blavatnik School of Government and the University of Oxford. This dataset includes 23 indicators grouped into five categories: containment and closure policies (C), economic policies (E), health system policies (H), vaccination policies (V), and miscellaneous policies (M), with most indicators representing the level of strictness of the policy. Four indicators (E3, E4, H4, and H5) are recorded as a US dollar value of fiscal spending, while V1 records categorical data and the ranked order of prioritized groups for vaccination. Additionally, ten indicators have a flag indicating whether they are targeted to a specific geographical region(flag = 0) or a general policy applied across the whole country/territory (flag =1). E1 has a flag for income support, and H7 has a flag to describe whether vaccine policy is funded by the government or at cost to the individual. Several indices are also calculated to provide an overall impression of government activity, with all indices being simple averages of the individual component indicators. For example, the Stringency Index is the average of all closing policies (C1-C8). The numerical policies in the dataset were normalized to a range of 0-100, taking into account their flag value, using the following formula:

$$p_{j,t} = 100 * \frac{v_{i,t} - 0.5 * (F_j - f_{j,t})}{N_j} \tag{4.1}$$

Where $P_{j,t}$ represents the scaled value of the jth policy on day t, F_j indicates whether the policy has a flag variable or not, with $F_j = 0$ for policies without flag variables and vice versa. The flag value of the jth policy on day t is represented by $f_{j,t}$, while $v_{i,t}$ represents the recorded value of the jth policy on day t. Finally, N_j represents the maximum value (range) of the jth policy. Furthermore, the daily number of confirmed cases was subjected to a simple moving average with a window of 7 to account for weekly periodicity in the data. In our counterfactual reasoning methods, the policy scales were first transformed to a range of 0-1 through the use of a min-max scaler, while the number of daily confirmed cases was scaled using a logarithmic function. Among the policies, we just have used the numerical ones and the policies that there were enough records for them in the dataset. The selected policies with their description can be seen in the Table 4.1.

Policies	Description
V1_Vaccine_Prioritisation	Reports the existence of a prioritised plan for vaccine (0: no plan, 1: a prioritised plan is in place, 2: universal/general eligibility; no prioritisation between groups)
V2A_Vaccine_Availability	Reports whether any categories of people are receiving vaccines (0 :no categories are receiving vaccines, 1 :vaccines are available to some categories, 2: vaccines are available to anyone over the age of 16 yrs, 3: vaccines are available to anyone over the age of 16 yrs PLUS one or both of 5-15 yrs and 0-4 yrs)
V2D_Medically_clinically vulnerable (Non-elderly)	Reports the number of categories selected from thematic group: V2_At risk age ranges below 60 (one or more selected counts as 1 x category), V2_Clinically vulnerable/chronic illness/significant underlying health condition (excluding elderly and disabled) , V2_Disabled people , V2_Pregnant people , V2_People living with a vulnerable/shielding person or other priority group (0: no categories are receiving vaccines, 1: 1 or 2 categories in group selected, 2: 3 or more categories selected or all from V2_General 16-19 years up to V2_General 80+ years present)

V2E_Education	Reports the number of categories selected from thematic group: V2_Educators, V2_Primary and secondary school students, V2_Tertiary education students (0: no categories are receiving vaccines, 1:1 category in group selected, 2: 2 or more categories selected or all from V2_General 16-19 years up to, V2_General 80+ years present)
V2F_Frontline workers (non healthcare)	Reports the number of categories selected from thematic group: V2_Police/first responders, V2_Airport/Border/Airline staff, V2_Factory workers, V2_Frontline retail workers, V2_Military, V2_Other high contact professions/groups (taxi drivers, security guards), V2_Frontline/essential workers (when subcategories not specified) (triggers an automatic 2) (0: no categories are receiving vaccines , 1: 1 or 2 categories in group selected, 2: 3 or more categories selected or all from V2_General 16-19 years up to V2_General 80+ years present)
V2G_Frontline workers (healthcare)	Reports the number of categories selected from thematic group: V2_Staff working in an elderly care home, V2_Healthcare workers/carers (excluding care home staff) (0: no categories are receiving vaccines, 1:1 category in group selected, 2: 2 categories selected or all from V2_General 16-19 years up to V2_General 80+ years present)

H1_Public information campaigns	Record on the availability of the public campaigns (0: no Covid-19 public information campaign, 1: public officials urging caution about Covid-19, 2: coordinated public information campaign (eg: social media))
H7_Vaccination policy	Record on vaccine delivery for different groups (0: No availability, 1: Availability for ONE of following: key workers/ clinically vulnerable groups (non elderly) / elderly groups, 2: Availability for TWO of following: key workers/ clinically vulnerable groups (non elderly) / elderly groups, 3: Availability for ALL of following: key workers/ clinically vulnerable groups (non elderly) / elderly groups , 4: Availability for all three plus partial additional availability(select broad groups/ages) , 5: Universal availability)
H8_Protection of elderly people	Record policies for protecting elderly people (0: no measures, 1: Recommended isolation, hygiene, and visitor restriction measures in LTCFs and/or elderly people to stay at home, 2: Narrow restrictions for isolation, hygiene in LTCFs, some limitations on external visitors and/or restrictions protecting elderly people at home, 3: Extensive restrictions for isolation and hygiene in LTCFs, all non-essential external visitors prohibited, and/or all elderly people required to stay at home and not leave the home with minimal exceptions, and receive no external visitors)

H6_Facial Coverings	Record on using facial coverings outside home (0: no policy, 1:recommended, 2: Required in some specified shared/public spaces outside the home with other people present, or some situations when social distancing not possible , 3: Required in all shared/public spaces outside the home with other people present or all situations when social distancing not possible, 4: Required outside the home at all times regardless of location or presence of other people)
H2_Testing policy	Record on who have access to testing (0: no testing policy, 1: only those who both (a) have symptoms AND (b) meet specific criteria, 2: testing of anyone showing Covid-19 symptoms, 3: open public testing (eg ”drive through” testing available to asymptomatic people)
H3_Contact tracing	Record on contact tracing after a positive diagnosis (0: no contact tracing, 1: limited contact tracing; not done for all cases, 2: comprehensive contact tracing; done for all identified cases)
E1_Income support	Record on the direct payment of the government to people who lost their job or cannot work (0: no measure, 1:government is replacing less than 50% of the lost salary, 2:overment is replacing 50% or higher percentage of the lost salary, 3:)
E2_Debt contract relief	Record if the government is freezing financial obligations for households (e.g: stopping loan repayments, preventing services like water from stopping, or banning evictions) (0: no measure, 1:narrow debt relief, 2: broad debt relief)

C1_School closing	Record on closing schools and universities (0:no measures, 1:recombed closing or all school opens, 2: require closing (some levels), 3: require closing all levels)
C2_Workplace closing	Record on closing workplaces (0:no measures, 1:recommend closing or all open or recommend working online, 2:require closing for some sectors, 3:require closing for all essential workplaces)
C3_Cancel public events	Record on cancelling public events (0: no measure, 1:recommend cancelling, 2: require cancelling)
C4_Restrictions on gatherings	Record on the limits on gatherings (0: no measure, 1: restriction on very large gatherings(above1000), 2: restriction on gatherings between 101-1000), 3: restriction on gatherings between 11-100, 4: restriction on gatherings of 10 people or less
C5_Close public transport	Record on the closure of public transportation (0: no measure, 1: require closing (reduce the volume/ type of transport), 2: require closing (or prohibit all people from using them))
C6_Stay at home requirements	Record on staying at home (0:no measure, 1:recommend not leaving home, 2:require not leaving home with exceptions on emergency)
C7_Restrictions on internal movement	Record of the restrictions on the movement between cities and regions (0: no measure, 1:recommend not to travel between cities, 2: internal movement restrictions in place)

C8_International travel controls	Record on the international travel restrictions (just for foreign travellers) (0: no measure, 1:monitoring arrivals, 2: quarantine arrivals for some or all regions 3:ban arrivals from some regions, 4:ban on all arrivals or border closures)

Table 4.1: Description of governments' policies during COVID 19 (continued)[1]

Table 4.2 presents the outcomes of the four methods applied to measure the impact of policies. The mutual information method provides non-negative values, where greater values indicate stronger dependency and effect of policies on the target variable (i.e., daily confirmed cases). The HSIC method determines whether a policy had a significant effect on the target variable or not. The counterfactual reasoning approach calculates the average error between the actual number of daily confirmed cases (scaled by logarithmic function) and the counterfactual value obtained through the method. Negative values indicate that a policy had a detrimental impact on controlling the disease, and its implementation at a lower level of strictness or omission would have resulted in a lower number of daily confirmed cases. Positive values indicate that a policy was effective in controlling the disease, and increasing its level of strictness could further reduce the number of daily confirmed cases. The counterfactual generation method proposes a new set of policy levels for reducing the number of daily confirmed cases on peak days. As the policies were trained on a range of 0-100 for the regressor, the results indicate the degree of increase or decrease in the level of each policy to achieve a lower number of COVID-19 cases in range of 0-100.So, policies with a positive effect are those for which the algorithm suggests an increase in their level, and vice versa.

Among the five methods used in our study, the causality-based methods were able to not only show the amount of dependency between each policy and the number of daily confirmed cases but also clarify the positive or negative effect of each policy on controlling the disease. On the other hand, the dependency-based methods can only show the amount of dependency between the policies and did not provide information on the positivity or negativity of their effect on controlling the disease.

According to our analysis, three of the methods (MI, Count_Res_GRU, and Count_Gen) have identified V2D_Medically_clinically vulnerable (Non-elderly) as the most effective vaccination policy for controlling daily confirmed COVID-19 cases, indicating that increasing its level could lead to a decrease in the number of cases. Additionally, V2F_Frontline workers (non healthcare) has been recognized as the second most effective vaccination policy, while V2E_Education has been ranked third among the vaccination policies.

All three causality-based methods recommend a lower level of V2G_Frontline workers (healthcare) for reducing the number of daily confirmed cases, whereas the dependency factor indicates a dependency of only 0.2 between the policy and the number of cases. Furthermore, four methods (MI, Count_Res_SVR, Count_Res_GRU, and Count_Gen) have determined that V1_Vaccine_Prioritisation and V2A_Vaccine_Availability have no effect on controlling the spread of COVID-19. However, since the value of these policies was zero for a significant portion of the dataset, their classification as ineffective may be due to poor data quality rather than an accurate representation of their effect. Also, the HSIC method has been identified this policy as having effect on the number of daily cases.

For economic policies, MI method recognizes E1_Income support as the most effective one while the three causality based methods recognize E2_Debt contract relief as the most effective one, and on the positivity or the negativity of the effects counterfactual generation methods considered both economic factors as policies that had positive impact on controlling the disease and suggests an increase in their level, while the other two causality methods consider those as having a negative effect on controlling the disease.

Four algorithms (methods (MI, Count_Res_SVR, Count_Res_GRU, and Count_Gen)) identified H7_Vaccination policy as one of the top four most effective health policies in reducing the number of daily confirmed cases. In addition, all three causality-based algorithms included H1_Public information campaigns among the top four most effective health policies. Count_Res_SVR and Count_Gen identified H8_Protection of elderly people as the least effective policy, while the other two methods ranked it as the third most effective policy. Almost all algorithms found that H1 and H7 had a positive impact on controlling the spread of the disease and recommended increasing their level. On the other hand, H8 had a positive effect but was considered the least effective policy by two algorithms. The results regarding the positivity or negativity of other policies varied among the three causality-based methods, and they provided different interpretations.

Four analyzed (MI, Count_Res_SVR, Count_Res_GRU, and Count_Gen) methods identified C1, C3, and C4 as among the top five most effective closure policies for controlling the spread of the disease, with three methods (Count_Res_GRU, Count_Res_SVR, Count_Gen) also recognizing C7 in this category. While all three causality-based methods indicate that C3 has a positive impact on reducing daily confirmed cases, and recommend an increase in its implementation, only the two counterfactual reasoning methods consider C1, C4, and C7 as policies with a positive effect, while the counterfactual generation method suggests reducing their implementation to decrease the number of daily confirmed cases. Interestingly, the counterfactual generation method considers C8 as the most effective closure policy with a positive effect, while the other methods do not assign much weight to it, and only one method concurs with its positive impact.

According to the findings obtained from the proposed algorithms, the vaccination policies were found to have the most significant effect in controlling the spread of the COVID-19 disease, which is consistent with our intuitive understanding. Furthermore, among the vaccination policies or vaccination availability, it was found that vaccinating people with disabilities and medically/clinically vulnerable individuals was the most critical aspect, followed by vaccinating frontline workers, and then those associated with schools and education were identified as the most important groups to be vaccinated.

Additionally, among the closure policies, it was found that school closures, cancellation of public events, and restrictions on gatherings were the most effective measures in controlling the spread of the disease, which is also consistent with our understanding. However, the algorithms also produced some counter-intuitive conclusions. For instance, the testing policy was found to be destructive, and the algorithms suggested a lower level of implementation. Moreover, for some policies, the algorithms provided different statements on their positivity or negativity.

A tabulated summary of the results obtained from all the algorithms for all policies can be found in Table 4.2, and these counter-intuitive findings can be observed by assessing the table results.

Policy/Method	Dependency (HSIC)	Dependency (MI)	Counterfactual Reasoning (GRU)	Counterfactual Reasoning (SVR)	Counterfactual Generation
V1_Vaccine_Prioritisation	Dependent	0.00056	4.2661e-10	0.0	0.0
V2A_Vaccine_Availability	Dependent	0.000501	4.3791e-10	0.0	0.0
V2D_Medically_clinically vulnerable (Non-elderly)	Dependent	0.31662	-0.060943	0.07960	55.32
V2E_Education	Dependent	0.27588	0.056011	0.03858	7.62
V2F_Frontline workers (non healthcare)	Dependent	0.26836	0.074795	0.07701	34.25
V2G_Frontline workers (healthcare)	Dependent	0.21252	-0.047364	-0.3603	-9.62
H1_Public information campaigns	Dependent	0.011948	0.52995	-0.73849	33.02
H7_Vaccination policy	Dependent	0.477148	0.205889	0.320958	24.24
H8_Protection of elderly people	Dependent	0.26528	0.12542	0.011315	1.05
H6_Facial Coverings	Dependent	0.365488	0.044294	0.24361	-6.23
H2_Testing policy	Dependent	0.23562	-0.03172	-0.4340	-12.05
H3_Contact tracing	Dependent	0.15982	-0.054270	-0.1251	37.43
E1_Income support	Dependent	0.42720	0.0152382	-0.04649	3.55
E2_Debt contract relief	Dependent	0.18861	-0.058103	-0.14097	20.32
C1_School closing	Dependent	0.472582	0.180043	0.60221	-5.07
C2_Workplace closing	Dependent	0.454668	0.110302	0.02218	1.01
C3_Cancel public events	Dependent	0.41243	0.047841	0.23158	12.55
C4_Restrictions on gatherings	Dependent	0.42311	0.06925	0.12588	-16.90
C5_Close public transport	Dependent	0.15011	0.014426	0.03298	-3.67
C6_Stay at home requirements	Dependent	0.378284	0.026564	0.00775	1.04
C7_Restrictions on internal movement	Dependent	0.17138	0.09443	0.2676	-18.19
C8_International travel controls	Dependent	0.071675	0.01698	-0.29042	52.41

Table 4.2: The results of the effect of policies on the number of daily confirmed cases obtained from all four methods.

Chapter 5

Limitations of the Proposed Methodology

Since in counterfactual reasoning, we are asking on situations that have not been seen in the dataset, the model must have the generalizability to be able to make a correct decision and bring a true output based on our counterfactual question(input). But it is more likely that our models are not generalized in a way to be able to answer counterfactual questions. We were able to show the low mean square error on both train and test data and also the predictions on both train and test data are following the trend, but these factors cannot represent if the model is generalized or not. They can only tell us that the model has not been overfitted on the data. Consider the case which is depicted in Figure 5.1, based on the figure the true underlying model is the complicated curve, but our dataset is just covering a small part of that. So, by training a deep learning/machine learning model on this data although we arrive at a good model based on the deep learning factors the resulted model is not generalizable. For this problem, we have tried to make sure that the distribution of the created counterfactual questions is close to the distribution of the dataset as it was explained in the methodology part, but as it can be seen in the results, there are some policies that the prediction of the models are different for them. Some of the models have recognized them as effective and some not. Although other available works in the literature stated confronting with these cases in their results as well, but we are discussing this to attract the users' attention in using the results carefully and by considering the shortages of the proposed methods.

The generalization problem may has been happened due to the dataset and ignoring some confounders that are affecting the COVID19 pandemic but are not recognized to us or have been ignored.

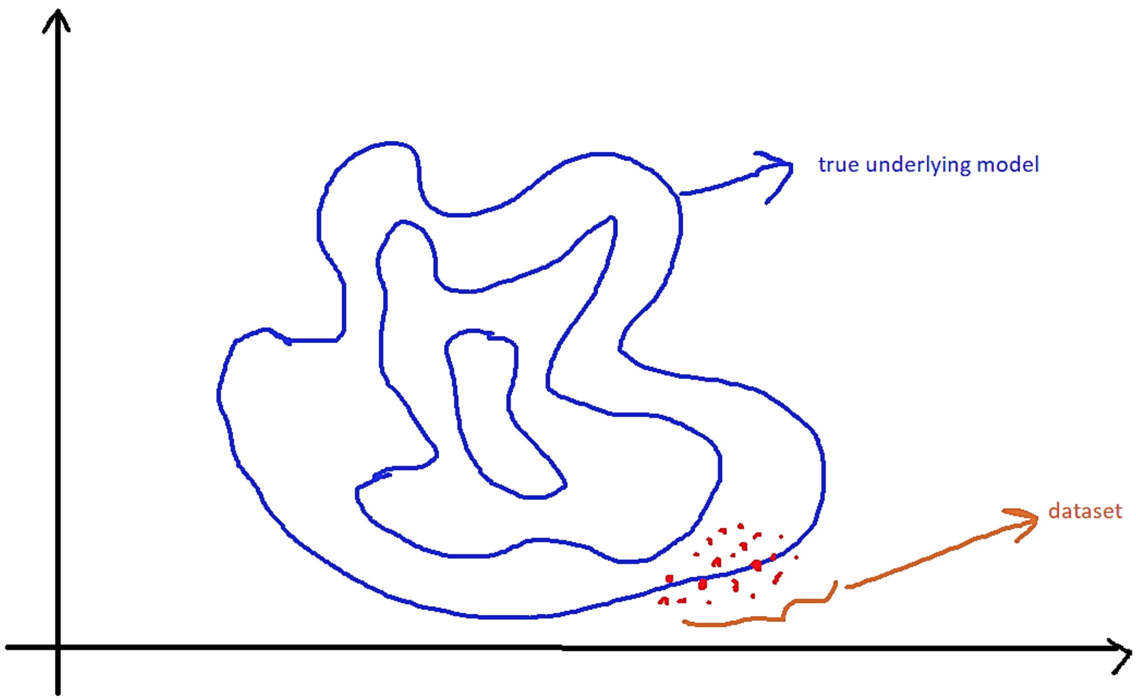

Figure 5.1: An illustration of the case when the dataset is not a good representative of the underlying model.

Chapter 6

Conclusion

This paper presented a study on the impact of government policies on the spread of COVID-19. By employing deep learning/machine learning models and counterfactual reasoning, we attempted to investigate the effect of policies on the number of daily confirmed cases. Our results provide insights into the impact of different policies on the spread of COVID-19 and can assist policymakers in addressing future pandemics.

References

[1] Blavatnik school of government, university of oxford. coronavirus government response tracker. https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker.

[2] Introducing gated recurrent units. https://subscription.packtpub.com/book/data/9781789956177/10/ch10lvl1sec43/introducing-gated-recurrent-units.

[3] An introduction to support vector regression (svr). https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2.

[4] Lecture 2: Entropy and mutual information - mcgill university. https://www.info612.ece.mcgill.ca/lecture_02.pdf.

[5] measuring statistical dependence with hilbert-schmidt norm. https://wiki.math.uwaterloo.ca/statwiki/index.php?title=measuring_Statistical_Dependence_with_Hilbert-Schmidt_Norm.

[6] Who director-general's opening remarks. https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[7] Martin C. J. Bootsma and Neil M. Ferguson. The effect of public health measures on the 1918 influenza pandemic in u.s. cities. *Proceedings of the National Academy of Sciences*, 2007.

[8] Hye Won Chung, CatherineApio, TaewanGoo, Gyujin Heo, Kyulhee Han, Taehyun Kim, Hakyong Kim, Yeonghyeon Ko, Doeun Lee, Jisun Lim, Seungyeoun Lee, and Taesung Park. Efects of government policies on the spread of covid-19 worldwide. *Nature*, 2021.

[9] Susanne Dandl, Christoph Molnar, Martin Binder, and Bernd Bischl. Multi-objective counterfactual explanations. *Parallel Problem Solving from Nature – PPSN XVI*, 2020.

[10] Rahul Dey and Fathi M. Salem. Gate-variants of gated recurrent unit (gru) neural networks. *IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2017.

[11] J. C. Gower. A general coefficient of similarity and some of its properties. *Biometrics*, 1971.

[12] Amir-Hossein Karimi, Gilles Barthe, Borja Balle, and Isabel Valera. Model-agnostic counterfactual explanations for consequential decisions. arXiv, 2019.

[13] Atalanti Mastakouri and Bernhard Schölkopf. Causal analysis of covid-19 spread in germany. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 3153–3163. Curran Associates, Inc., 2020.

[14] Atalanti A. Mastakouri and Bernhard Schölkopf. Necessary and sufficient conditions for causal feature selection in time series with latent common causes. *arXiv preprint arXiv:2005.08543*, 2020.

[15] Judea Pearl. The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 2019.

[16] Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell. *Causal inference in statistics: A Primer*. Wiley, 2021.

[17] Jonas Peters, Dominik Janzing, and Bernhard Scholkopf. *Elements of causal inference foundations and learning algorithms*. The MIT Press, 2017.

[18] Statistics and Computing. A tutorial on support vector regression. 2004.

[19] Chernozhukov Victor, Kasahara Hiroyuki, and Shrimpf Paul. The association of opening k-12 schools and colleges with the spread of covid-19 in the united states: County-level panel data analysis. *SSRN Electronic Journal*, 2021.

[20] Julius von Kügelgen, Luigi Gresele, and Bernhard Schölkopf. Simpson's paradox in covid-19 case fatality rates: A mediation analysis of age-related causal effects. *IEEE Transactions on Artificial Intelligence*, 2021.