Analyzing Covid-19 Mortality Rate and Policy Effectiveness using OxCGRT and OWID

Abhijeet Sahdev, Bhaskar Kalli, Subodh Dahal
New Jersey Institute of Technology

Abstract—**Given the immense research conducted on datasets about the pandemic that started in the year 2019, there is a lack of justification for selecting lagged features when considering the impact of policies on Covid-19. The project proposes three sets of lagged Policy features from OxCGRT and compares the predictive power of machine learning models for regression tasks using multivariate linear regression and classification tasks using decision trees. The linear models were evaluated using a combination of coefficient of determination (r^2) and mean squared error, whereas the decision trees were compared using weighted F1 scores. We achieved an r^2 score of 0.6915708090641737 with mse of 0.0047251403134646 for predicting daily mortality rates and a weighted F1 score of 0.7152737209142526 for classifying effective policies, considering lag effects after 21 days in both the cases globally. For classification tasks, lag effect of 21 days was the best.**

1. Introduction

The unprecedented speed of Covid-19 spreading and multiplying caught a lot of nations by surprise and unable to react in time. This necessitated rapid implementation of various untested policies to hopefully curb the spread and reduce the risk of fatality. These policies created massive economical burdens , the after affects of which are still prevalent to this day, especially for countries that were not able to manage their economic policies well, or for countries that have low GDP and the economy loss would have major effects.

1.1. Policy Responses to COVID-19

During the rapid spread of the virus, the government implement various strategies in several severity levels based on how the population was affected, it started with containment policies along with healthcare policies to mitigate the impact caused by the virus and when there was not much of an improvement, stricter containment policies, i.e lock downs were announced which was a drastic step due to the effect it would have on the economy, to combat this, several economic policies were put into place, in hopes to control the damages caused by the virus, until vaccines were developed and ready for general public, which helped reduce the risk of a fatality from the virus

1.2. The Importance of Data in Policy Making

The project believes that data of such precedents are necessary to prevent future tragedies caused from any similar pandemics and would help in deciding the most effective policy as per the current situation such that the after affects due to implementing the policy would be low and the pandemic could be contained from becoming too dangerous, the project hopes to use the data gathered by oxcgrt and owid and study how implement certain policies affected the mortality factor and further examine what factors influenced the decision of implementing such policy

1.3. Outcome of the Study

In order to examine their effects on the mortality rate and the efficacy of the policies, the project creatively established lags of 7, 14, and 21 days in the major policy measures. Rather than concentrating only on specific nations, this strategy offered insight into global trends. The suggested approach demonstrated performance that was almost on par with the different algorithms suggested in other research, underscoring the need of adding lag information to the dataset to enhance and help anticipate how effective a policy was against the infection. This improves the algorithm's robustness and global applicability and makes a significant contribution to the assessment of policies globally.

Keywords: Artificial Intelligence, Machine Learning, COVID-19, non-pharmaceutical interventions.

2. Objectives

The main objectives of the project are:

- Examine the current methods used to analyze the efficacy of different COVID-19 policies.
- To prepare the data for use in training different machine learning models, clean and preprocess it to make it comprehensive.
- To conduct time series analysis and geospatial analysis using the given dataset.
- To append features from other sources to create meaningful distributions.
- Create and put into use machine learning models that, using a variety of candidate features to forecast the efficacy of policies.

- After evaluating these models, decide which of the implemented models performs the best.
- Describe the future work's scope.

3. Methodology

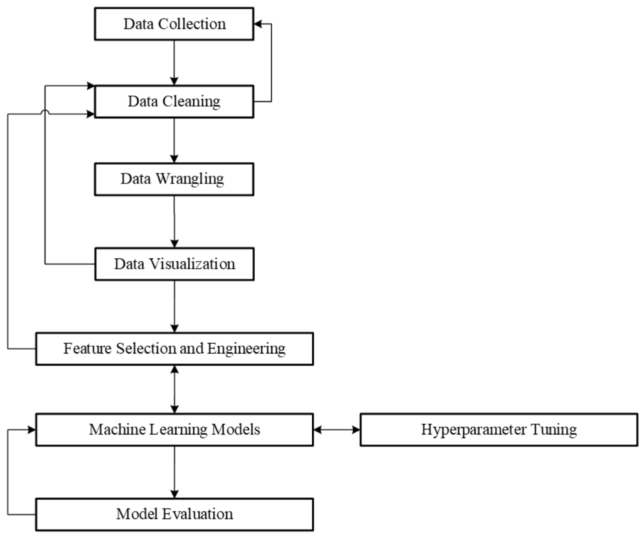


Figure 1: Workflow

Figure 1 captures the overview of the key steps and their interactions through out the flow of the project.

3.1. Preliminaries

3.1.1. Multivariate Linear Regression. Multivariate Linear regression is a linear model to learn the relationship between a dependent variable and $>=2$ independent variables. Mathematically, it is defined as:

$$\alpha = \theta_0 + \theta_1\beta_1 + \theta_2\beta_2 + \ldots + \theta_n\beta_n \tag{1}$$

where α is the target variable, θ_0 is the bias term, and $\theta_1, \theta_2, \ldots, \theta_n$ are coefficients or weights of $\beta_1, \beta_2, \ldots, \beta_n$, the independent features.
For the scope of the project, a model learns patterns on train dataset and predicts on test dataset. Train and test datasets are obtained by splitting the given dataset. So let n be the size of the dataset with m independent features, so now,

$$\alpha = \begin{bmatrix} 1 & \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,n} \\ 1 & \beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \beta_{m,1} & \beta_{m,2} & \cdots & \beta_{m,n} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}$$

Considering this, the algorithm to predict the daily mortality rate using multivariate linear regression with gradient descent is as follows: where $mepochs$ is the maximum number of iterations through the entire dataset and c is termed as the learning rate. Polynomial features in this context are features that help the model capture relationships between independent features by considering their combinations, increasing the degree of the feature set and thus, capturing patterns in higher dimensions.

Algorithm 1: MVRDMR - Multivariate Regression for Daily Mortality Rate

Data: β (feature matrix), α (target variable)
Result: θ (weight vector including bias)
k = 0; Initialize a weight vector $\theta = (\theta_0, \theta_1, ..., \theta_n)$ randomly;
while *k!= mepochs* **do**
    $\theta_0 \leftarrow \theta_0 - c\sum_{m=1}^{n}(\theta_0 + \sum_{i=1}^{d}\theta_i\beta_{m,i} - \alpha_m)$;
    for $i \leftarrow 1$ **to** d **do**
        $\theta_i \leftarrow \theta_i - c\sum_{m=1}^{n}(\theta_0 + \sum_{j=1}^{d}\theta_j\beta_{m,j} - \alpha_m)\beta_{m,i}$;
    end
    k++;
end
return θ;

3.1.2. Elastic Net. It is a regularisation technique that aims to reduce overfitting by penalising the weights of the features, proportional to their tendency to learn outliers (L_2), and also reduces the dimensions of the feature set (L_1). Such penalisation is carried out by adding terms to the loss function such as (6).

3.1.3. Decision Trees. Given the importance of interpretation in terms of policy making, the project utilises this supervised learning technique for its classification tasks since the algorithm has an intuitive approach of 'asking questions' to the dataset and dividing it into different nodes. The question to be asked are decided by calculating Gini Index for the q^{th} node, which is given by the following equation:

$$Gini_q = 1 - \sum_{l=1}^{Z}(z_{l,q}^2) \tag{2}$$

where $Gini_q$ is the Gini impurity of the q^{th} node,Z is the total number of classes and $z_{l,q}$ is the proportion of class l instances among the training instances in the q^{th} node. For a split, the index is calculated for all possible feautres, and the one that introduces homogeneity in the tree, i.e, Gini = 0, is considered the best split.

3.2. Performance Evaluation Parameters

3.2.1. Mortality Rate. Mortality rate (M) is defined as the number of deaths reported on that day by the population in concern, times 1000 or per 1000 people.

$$M(\text{per thousand}) = \frac{\text{Difference in cumulative deaths}}{\text{Population of the country}} \times 1000 \tag{3}$$

Where: Difference in cumulative deaths is total number of deaths on that day, and the difference is calculated with reference to the previous day.

3.2.2. Policy Effectiveness. A policy is said to effective if the number of deaths per 1000 people of the population of a country reduces the next day.

$$P.E = \begin{cases} 0 & \text{if diff(daily mortality rate)} > 0 \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

3.2.3. For Regression Models.

1) Coefficient of Determination (R^2) The project uses the sum of squares variant to define the metric, which is given as follows:

$$R^2 = 1 - \frac{\sum(\alpha_i - \hat{\alpha}_i)^2}{\sum(\alpha_i - \bar{\alpha})^2} \tag{5}$$

where the summation is carried for the total number of tuples in the dataset ($i = 1$ to size of dataset) and carry the following meaning:

- a_i : actual value from the dataset.
- $(\hat{a}_i)$: predicted value for the dataset.
- $\overline{a}$: expected value of given dataset.

2) Mean Square Error For a given size t of a dataset having corresponding predicted $\hat{\alpha}_i$ and true values α_i, where ($i = 1$ to t), mean squared error is given by:

$$\text{MSE} = \frac{1}{t}\sum_{i=1}^{n}(\hat{\alpha}_i - \alpha_i)^2 \tag{6}$$

3) Root Mean Square Error It is the square root of eq(6)

$$\text{RMSE} = \sqrt{\frac{1}{t}\sum_{i=1}^{n}(\hat{\alpha}_i - \alpha_i)^2} \tag{7}$$

3.2.4. For Classification Models. The metrics defined in this section use the following definitions for the terms that are reused multiple times:

- a (True Positives) : situations where the predicted value of a model is positive for the positive class in the dataset.
- b (True Negatives): Situations where the predicted value of a model is negative for the negative class in the dataset.
- c (False Positives): Situations where the predicted value of a model is positive but the actual value is negative in the dataset.
- d (False Negatives): Situations where the predicted value of a model is negative but the actual value is positive in the dataset.

$$\text{Accuracy} = \frac{a + b}{a + b + c + d} \tag{8}$$

$$\text{Precision} = \frac{a}{a + c} \tag{9}$$

$$\text{Recall} = \frac{a}{a + d} \tag{10}$$

Considering the imbalance present in the dataset, the project uses weighted F1 score to compare models which is defined as followed:

$$\text{Weighted F1 Score} = \frac{s_1}{S} \cdot \text{F1}_1 + \frac{s_2}{S} \cdot \text{F1}_2$$

where:

- s_1 and s_2 are the number of values in our classes (ineffective - 0 and effective - 1), respectively.
- S is the size of the dataset.
- F1_1 and F1_2 are the F1 scores for each class, computed as:

$$\text{F1}_1 = 2 \cdot \frac{\left(\frac{a}{a+c}\right) \cdot \left(\frac{a}{a+d}\right)}{\frac{a}{a+c} + \frac{a}{a+d}}$$

$$\text{F1}_2 = 2 \cdot \frac{\left(\frac{b}{b+d}\right) \cdot \left(\frac{b}{b+c}\right)}{\frac{b}{b+d} + \frac{b}{b+c}}$$

3.3. Data Pre-processing

Figure 2 shows the extensive workflow that has been performed on the datasets to get it to the standard that the project required for creating models. OxCGRT provides daily level metrics for policies implemented by different countries. This dataset containing about 24 different policies has measurements ranging from 0 to 5 for ordinal scale or continuous values. However, due to not having a few necessary features to understand the interaction between how the government chose a particular policy and why at the specific level for a country, this was supplemented by another dataset from OWID which contained a few additional metrics.

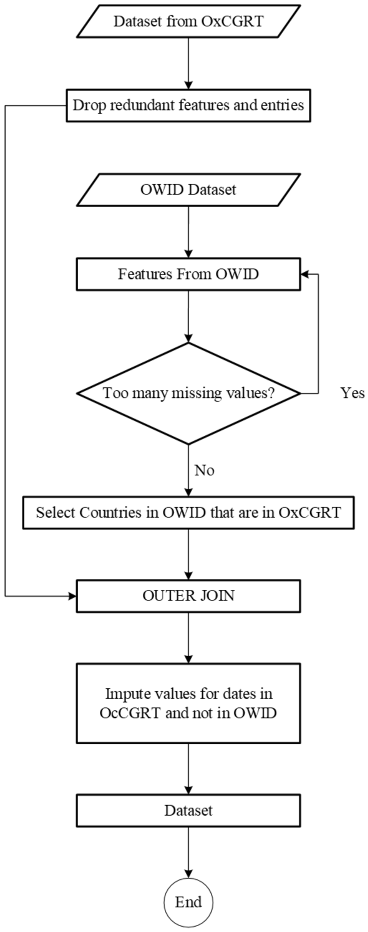


Figure 2: Data Collection

Initial data cleaning includes merging candidate features from a second dataset from OWID to get GDP, population, and a few others. This new dataset then has its columns converted into categorical types wherever possible along with converting the date column into a date object, as highlighted in Table 1. As there was a lot of missing data for quite a few columns, entries were dropped where the columns 'Confirmed Cases' and 'Confirmed Deaths' had missing values due to the possible impact of these on the remaining columns.

This meant that 21 out of the 184 initial countries had to be dropped, leaving 163 countries, which had their indexes aggregated and plotted over a world map (Figure 3) to understand the strength of the policy implemented over the period of three years. The main observation target was the government response index due to its weighted average of various policies implemented The calculation of which is shown in the Table 1.

The description of policy measures considered in the weighted indexes described in Table 1 are as follows:

- C1 : School Closure Policy

TABLE 1: Columns and Their Significance

Name	Description
CountryCode	ISO Country Code
Date	Date of the record in YYYYMMDD
ConfirmedCases	The cumulative number of reported covid-19 cases since the beginning of the pandemic
ConfirmedDeaths	The cumulative number of deaths attributed to covid-19 since the beginning of the pandemic
StringencyIndex_Average	Weighted Average of C1 to C8 and H1
GovernmentResponse Index_Average	Weighted Average of C1 to C8, E1, E2, and H1 to H8
ContainmentHealth Index_Average	Weighted Average of C1 to C8 and H1 to H8
EconomicSupportIndex	Weighted Average of E1 and E2
Population	Population of Country
GDP_Per_Capita	GDP of a Country
Life_Expectancy	Average life expectancy
Hospital_Beds_Per_Thousand	Number of hospital beds available per thousand people in country
Mortality Rate	Percentage of confirmed deaths over total population
V1	Vaccine Prioritization Policy
V3	Vaccine Funding Policy
H5	Vaccine Development Investment

- C2 : Workplace Closure Policy
- C3 : Public Event Cancellation Policy
- C4 : Gathering Size Limit Policy
- C5 : Public Transport Closure Policy
- C6 : Stay-at-Home Order Policy
- C7 : Restrictions on Internal Movement Policy
- C8 : International Travel Restriction Policy
- H1 : Public Information Campaign Policy
- H2 : Covid-19 Testing Policy
- H3 : Contact Tracing Policy
- H6 : Facial Covering Policy
- H7 : Vaccination Rollout Policy
- H8 : Elderly Care Protection Policy
- E1 : Income Support Policy
- E2 : Debt/Contract Relief Policy

Subsequently, CountryCode and MajorityVaccinated columns were encoded (binary and label, respectively) whereas Date was convered to datetime and its attributes were extracted.

3.4. Target Feature and Lagged Features

For regression tasks, the project aimed to model daily mortality rates globally. Utilising Eq 3 and drawing insights from Figure 5, lagged features (7 Day, 14 Day and 21 day) in Figure 6 were proposed to capture Figure 4. The project attempts to measure the impact a policy had by the increase or decrease in mortality rate, after final cleaning and dropping of entries with missing values, the dataset had 149 total countries with entries for each day for three years totaling 1096 entries per country. With a total of 19 rows selected as the candidate features, out of which mortality

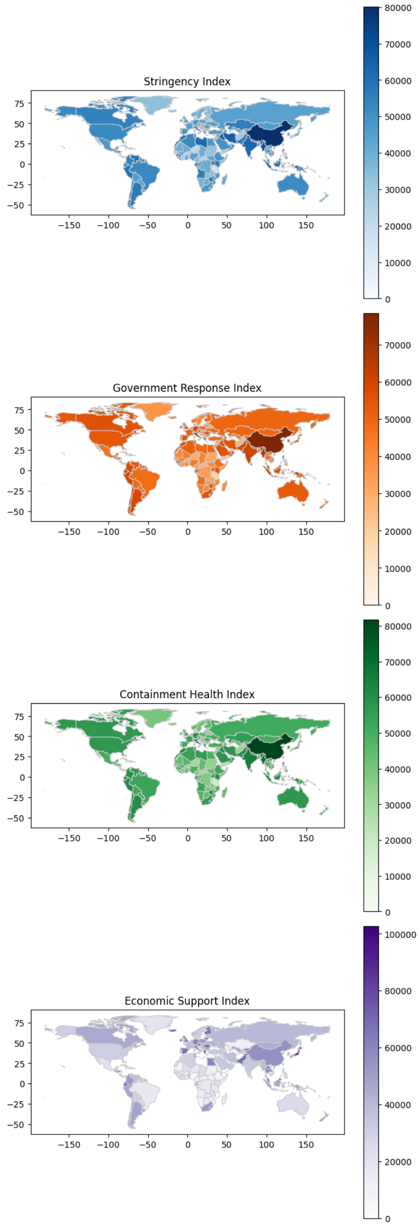


Figure 3: Mean Global Indexes for 3 years

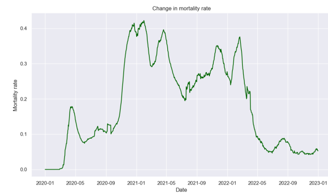


Figure 4: Change in mortality rate

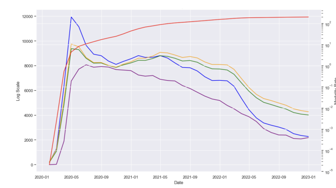


Figure 5: Mortality rate with indexes

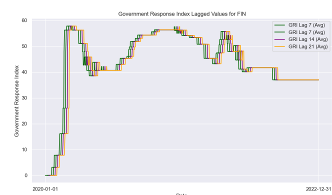


Figure 6: Lagged Features

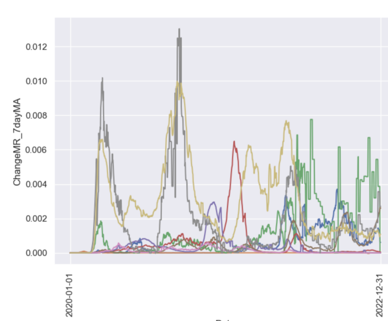


Figure 7: 7 day moving average

rate and confirmed cases had moving averages created for them to smooth out the fluctuations in the graph.

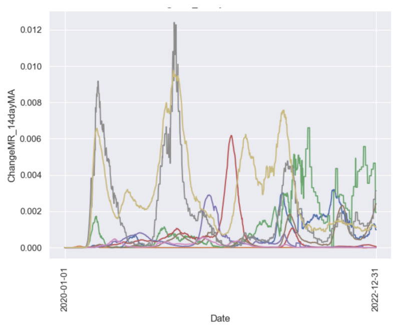


Figure 8: 14 day moving average

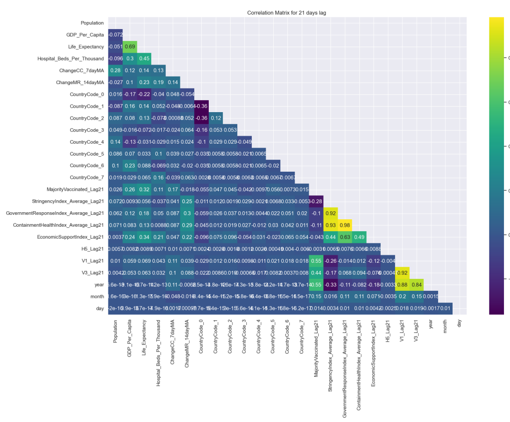


Figure 9: Correlation matrix

The project then applied lag to each of the indexes and created moving averages for confirmed cases and mortality rate which were adjusted to 7 day for confirmed cases and 14 day for mortality rate. Due to the 7 day moving average for mortality rate not being that smooth (**Figure 7**) compared to 14 day moving average (**Figure 8**) for mortality rate.

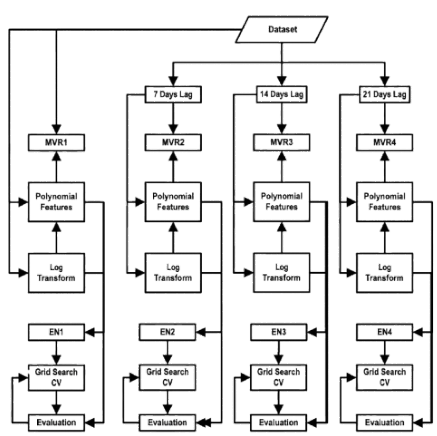


Figure 10: Regression workflow

3.5. Experimental Setup

The project performed various experiments, utilizing different combinations of features, aiming to predict the change in mortality rate using government policies only.

3.5.1. Regression Analysis of Daily Mortality Rate. The first model to be tested was MVLR to predict mortality rate (MVLR) based on the lagged features, where it was found that it was not able to properly predict the data, even after performing elastic net Grid Search (ENGS) hyper parameter tuning due to the non-linear nature of data.

This prompted the use of Polynomial Regression to gain insight into how the model interacts with the data provided to it, the score indicated that the model performed a bit better compared to LRMR and but still needed a bit more improvement. The project then utilized Grid search techniques combining polynomial and logarithmic features as shown in **Figure 10** to get the best model for set number of iterations, out of which the model that utilized data lagged by 21 days showed the highest R^2 and minimum MSE.

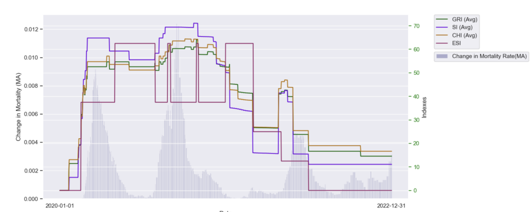


Figure 11: Change in indexes and mortality rate for Sweden

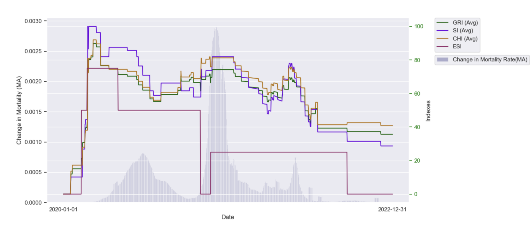


Figure 12: Change in indexes and mortality rate for India

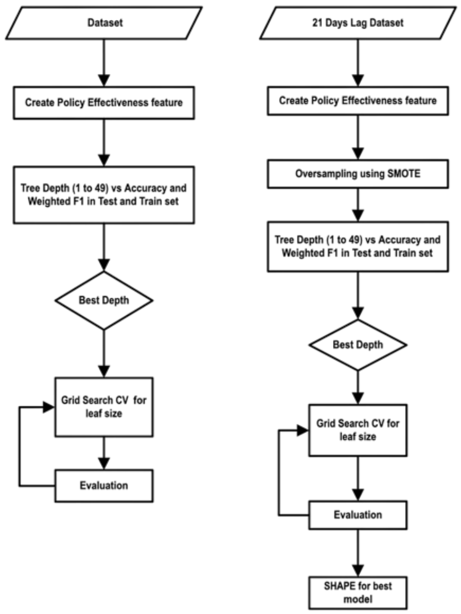


Figure 13: Classification workflow

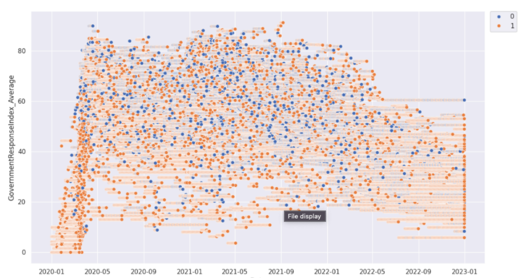


Figure 14: GRI with Time and Our Classes

3.5.2. Classification of Policy Effectiveness. Fig 14 clearly depicts that there is no linear decision boundary in the distribution. Moreover, it is noticeable that a majority of the labels belong to effective class during the initial periods (Jan, 2020) but that was the period when the virus spread exponentially and thus it needs to counteracted. Thus, it may not be appropriate to say that the policies were effective during this time but again, certain countries can be exceptions to this too and if we drop this entire period, we shall loose the initial peak in mortality rates and the associated first line of action that was implemented to combat such an event.

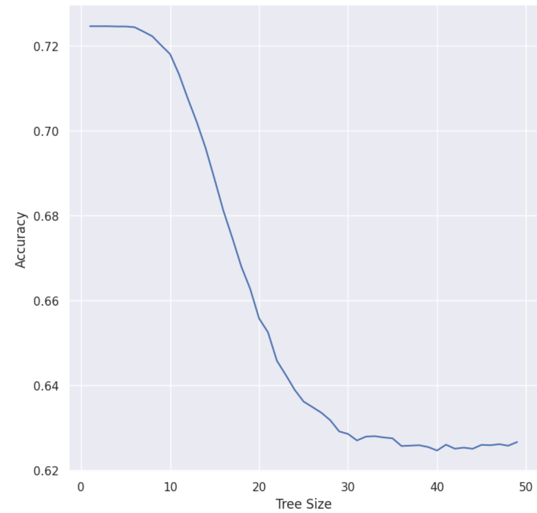
Given the highly correlated features in Figure 9, the initial experiment decided to experiment with collinear features since these features, such as GovernmentResponseIndex_Average and EconomicSupportIndex, where the latter could be assumed to be represented by the former, but the exact weights were not reported on [1]. Thus, justifying this experiment. Moreover, Fig 11 and Fig 12 show the same trend in GRI_Average but not for ESI.

First, the accuracy scores for each tree depth till 49, were recorded and plotted as shown in Tab 4 and Figure 15a. Indicating that the tree had overfit. Corresponding, weighted F1 scores on train (in blue) and test set (in yellow) were plotted to observe this trend of over-fitting. Fig 15b depicts that trend. Subsequently, a Grid Search Cross Validation tuning was conducted to find the best model to view its confusion matrix. The result was a tree with a weighted F1 score of 0.984516 (base) as reported in 5. Upon inspecting Fig 15c, it is clear that the tree predicts the positive class correctly (33202, True Positives) (i.e, effective policy) more better than it predicts the negative class (True negatives 2752). Moreover, it identifies a majority of the negative labels as members of a positive class too (10739, False positive). The false negative count is 2299, least here.

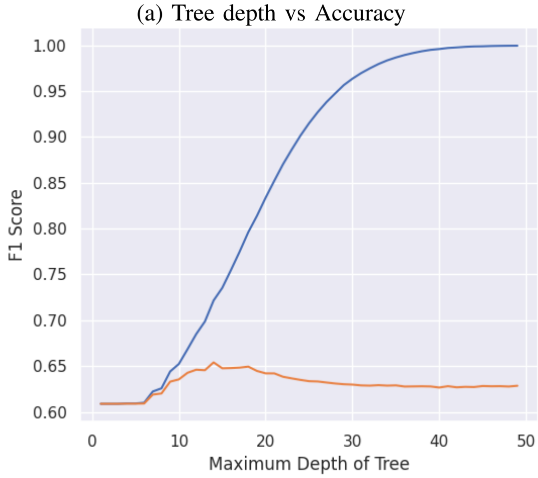
Before applying SMOTE, the models built on two datasets were compared on their weighted F1 scores (their distribution being 118334 for effective policies (1) and (0) for 44970) for varying tree depths, were compared on their accuracy measures. A peak was observed by a drop in the accuracy vs tree depth graph as depicted from the values in Table 4.

Now, a dataset with 21 days lag was tested too and it showed a similar trend. However, with this dataset, the goal was to iteratively eliminate collinear features (Fig 9 and arrive at a set of features that only represented policy indexes and other population data by viewing the features of the best model from Grid Search CV. Attributes of date (day, month and year) were still kept along with the binary encoded 'CountryCode'. The pipeline for 21 days lagged dataset was followed as shown in Fig 13. The exact same set of experiments were conducted with this dataset too, and its results are depicted in Tab 4 and 16. The maximum tree depth with the highest F1 score on test set was 14. Yet, the experiment proceeded with 13 to test if it would be lower than the results obtained with the dataset with no lags. The subsequent Grid Search CV resulted in the best model whose confusion matrix is shown in Fig 16c. The number of true positives has reduced, yet it still fairly high, true negatives have substantially increased because of oversampling done by SMOTE. The number of False negative (15,146) has increased. Comparing the values of precision and recall for the no lagged and the 21 day lagged dataset, it is clear that the model , which was imbalanced, was better at predicting the positive class, whereas the model built on SMOTE 21 Day lagged dataset will be considered a better judge for classifying positive labels since it takes equal number of negative labels into its learning algorithm.

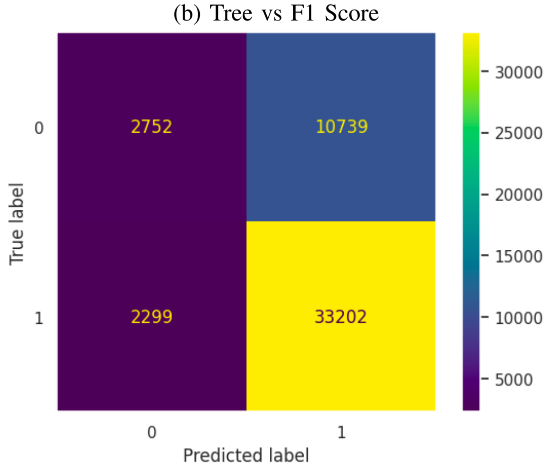
Taking the best model, SHAP values were calculated to see which features were important in decided the label of the test data points. Figure 17 indicates that life_expectancy plays an influential role in driving the model to classify a label as 0. Whereas Figire 18 indicates that GDP_Per_Capita would infuence the model to predict a label as 1 (effective).



(a) Tree depth vs Accuracy

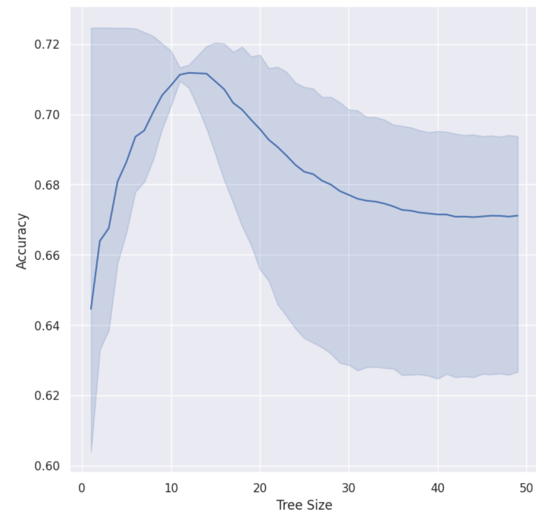


(b) Tree vs F1 Score

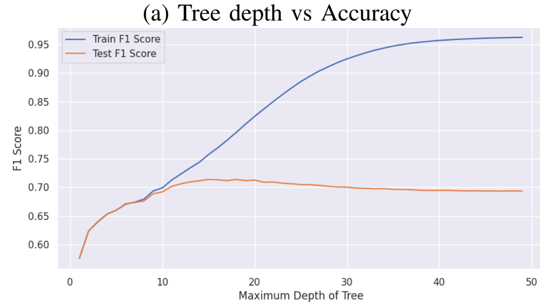


(c) Confusion Matrix

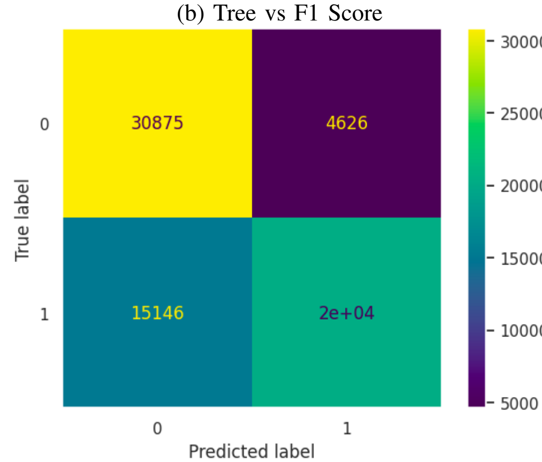
Figure 15: With no lagged features



(a) Tree depth vs Accuracy



(b) Tree vs F1 Score



(c) Confusion Matrix

Figure 16: With 21 days lagged features

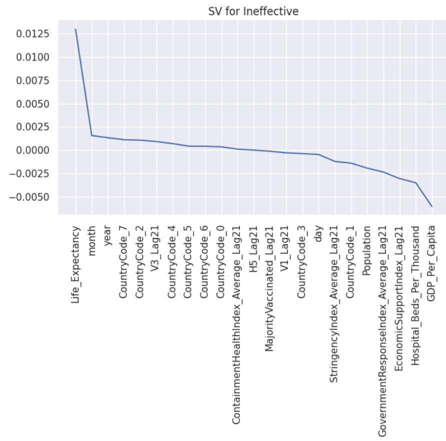


Figure 17: Mean SHAP Values for Features to predict - 0

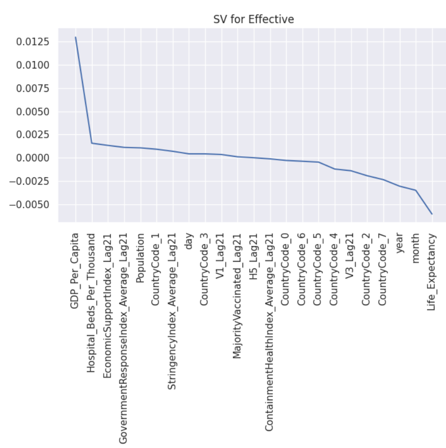


Figure 18: Mean SHAP Values for Features to predict - 1

4. Results

TABLE 2: Comparison Table

S.No.	Ref	Year	Methodology	Evaluation Parameters with their Corresponding Values
1	[2]	2020	Group-based trajectory modelling and multivariable linear regression	Time to a High-Level Stringency Index (SI¿80) from the Start of Response: Associated with 0.744 days earlier to the peak, SE=0.08, p¡0.001, R²=0.65. Time to a High-Level SI from the First Reported Case: Associated with 0.765 days earlier to the peak, SE=0.08, p¡0.001, R²=0.42.
2	[3]	2021	Multivariable linear regression	Analysis identified four trajectories based on initiation timing of responses. Earlier start to a high SI level decreased time to peak daily cases by 0.44 days (SE: 0.08, $P < 0.001$, $R^2 = 0.65$).
3	[4]	2022	Linear and non-linear regression	Earlier and stricter school closures reduced daily deaths per million by 1.23, whereas stricter workplace closures reduced it by 0.26.
4	[5]	2022	Random forest and Gradient boost	RF scored 79.67% and GB scored 80.10% in predicting policy effectiveness.
5	[6]	2023	KNN for clustering countries based on mortality rates	High Mortality Cluster includes Peru, Mexico with 37.93% mortality, Moderate Mortality Cluster includes USA, UK with 20.69% mortality.
6	[7]	2023	ES-LSTM model	Predicted growth rates with MAE of 0.0312, RMSE of 0.0384, and MAPE of 2.8460.

TABLE 3: Metrics for predicting Daily Mortality Rate

Index	Dataset and parameters	R^2	MSE	textbfRMSE
0	globalMR_NoLag	0.6878070282764837	0.0047828014974692	0.0691578014215988
1	globalMR_log_NoLag	0.8253162063994512	0.0027110717837841	0.0337679748188331
2	globalMR_logT_NoLag	0.9768742983017302	0.0003619386020523	0.0337679748188331
3	globalMR_lopo_NoLag	0.9228828318820816	0.0012069549450653	0.0347412571025482
4	globalMR_ENGS_NoLag.alpha(1e-06)_l1Ratio(1.0	0.6879123537430505	0.004781187909577	0.0691461344514431
5	globalMR_ENGS_NoLag.alpha(1e-06)__l1Ratio(1.0_2000	0.6878661414016861	0.0047818958834121	0.06915125366479
6	globalMR_ENGS_NoLag.alpha(5e-06)__l1Ratio(0.7_2000	0.6880202821966226	0.0047795344438807	0.0691341771042425
7	globalMR_Lag7	0.6839473126798008	0.0048419324043366	0.0695839953174337
8	globalMR_log_Lag7	0.8065294092800872	0.0030026406496084	0.0781828392289864
9	globalMR_logT_Lag7	0.9745597244563456	0.0003981638216316	0.0687396560470347

Continued on next page

TABLE 3 – continued from previous page

Index	Dataset and parameters	R^2	MSE	RMSE
10	globalMR_lopo_Lag7	0.9233129967111912	0.0012002224679742	0.0346442270511873
11	globalMR_lopo_Lag7_5e-06	0.9175273896402364	0.0012907725651175	0.0359273233781411
12	globalMR_ENGS_Lag7.alpha(1e-06)__l1Ratio(0.1	0.6839926330896081	0.0048412380949077	0.0695790061362461
13	globalMR_ENGS_Lag7.alpha(1e-06)__l1Ratio(0.1_2000	0.6839542245655944	0.0048418265141318	0.0695832344328129
14	globalMR_Lag14	0.6884833451731365	0.0047724403114111	0.069082851066029
15	globalMR_log_Lag14	0.778005286589566	0.0034453316548215	0.0687396560470347
16	globalMR_logT_Lag14	0.9765285259214253	0.0003673502593325	0.0687396560470347
17	globalMR_lopo_Lag14	0.9273625906496532	0.0011368425806051	0.0337170962659179
18	globalMR_ENGS_Lag14.alpha(1e-06)__l1Ratio(0.1	0.6885014116896306	0.0047721635320792	0.0690808477950237
19	globalMR_ENGS_Lag14.alpha(1e-06)__l1Ratio(0.1_2000	0.6884871068758807	0.0047723826820628	0.0690824339616288
20	globalMR_Lag21	0.6915708090641737	0.0047251403134646	0.0687396560470347
21	globalMR_poly2_Lag21	0.6915708090641737	0.0047251403134646	0.0687396560470347
22	globalMR_poly3_Lag21	0.6915708090641737	0.0047251403134646	0.0687396560470347
23	globalMR_poly4_Lag21	0.6915708090641737	0.0047251403134646	0.0687396560470347
24	globalMR_poly5_Lag21	0.6915708090641737	0.0047251403134646	0.0687396560470347
25	globalMR_poly6_Lag21	0.6915708090641737	0.0047251403134646	0.0687396560470347
26	globalMR_poly7_Lag21	0.6915708090641737	0.0047251403134646	0.0687396560470347
27	globalMR_log_Lag21	0.7643177220616741	0.0036577610348769	0.0687396560470347
28	globalMR_logT_Lag21	0.9775889591247132	0.0003507534997541	0.0687396560470347
29	globalMR_lopo_Lag21	0.9271432079002316	0.0011402761233653	0.0337679748188331
30	globalMR_ENGS_Lag21.alpha(1e-05)__l1Ratio(0.1	0.6915694293903878	0.0047251614500898	0.0687398097909056
31	globalMR_ENGS_Lag21.alpha(1e-05)__l1Ratio(0.1_2000	0.6915705443658691	0.004725144368647	0.0687396855437021

TABLE 4: Decision Tree metrics

Index	Dataset	Tree Size	Weighted F1	Accuracy
0	Updated_Base_	1.0	0.6089270533914964	0.7246285107772698
1	Updated_Base_	2.0	0.6089270533914964	0.7246285107772698
2	Updated_Base_	3.0	0.6089270533914964	0.7246285107772698
3	Updated_Base_	4.0	0.6090494154498843	0.7245672762900065
4	Updated_Base_	5.0	0.6090494154498843	0.7245672762900065

Continued on next page

TABLE 4 – continued from previous page

Index	Dataset	Tree Size	Weighted F1	Accuracy
5	Updated_Base_	6.0	0.6091975088508409	0.7244039843239712
6	Updated_Base_	7.0	0.6189790310758861	0.7233629980404964
7	Updated_Base_	8.0	0.6201997352894292	0.7222607772697583
8	Updated_Base_	9.0	0.6330986059354139	0.7201175702155453
9	Updated_Base_	10.0	0.6355367540908874	0.7180560091443501
10	Updated_Base_	11.0	0.6427293000119678	0.7133001306335728
11	Updated_Base_	12.0	0.6460896940099635	0.7075236773350752
12	Updated_Base_	13.0	0.645552625453041	0.7020329849771392
13	Updated_Base_	14.0	0.654070574321925	0.6958483017635533
14	Updated_Base_	15.0	0.6476358845081939	0.6885613977792293
15	Updated_Base_	16.0	0.6479032882797944	0.6810703788373612
16	Updated_Base_	17.0	0.6483709251780128	0.6747019921619857
17	Updated_Base_	18.0	0.6494500323184937	0.6680274330502939
18	Updated_Base_	19.0	0.6447359916946747	0.662740855649902
19	Updated_Base_	20.0	0.6421075994670477	0.6557805355976486
20	Updated_Base_	21.0	0.6420772520074732	0.6525146962769431
21	Updated_Base_	22.0	0.6383877932055504	0.6458197256694971
22	Updated_Base_	23.0	0.636627799850039	0.642472240365774
23	Updated_Base_	24.0	0.6350012916346419	0.6389614630960156
24	Updated_Base_	25.0	0.6335706314103953	0.6362059111691705
25	Updated_Base_	26.0	0.6332681283193758	0.6348995754408883
26	Updated_Base_	27.0	0.6321430982798201	0.6335728282168517
27	Updated_Base_	28.0	0.6310738129967017	0.6317970280862182
28	Updated_Base_	29.0	0.6302521188681788	0.6291639451338994
29	Updated_Base_	30.0	0.6298711618670546	0.6285720117570216
30	Updated_Base_	31.0	0.6289751280757877	0.6270411495754409
31	Updated_Base_	32.0	0.6287325919812031	0.6279392553886349
32	Updated_Base_	33.0	0.6292521118254515	0.6280413128674069
33	Updated_Base_	34.0	0.6287528022163668	0.6277555519268452
34	Updated_Base_	35.0	0.6290340194475194	0.6275310254735467
35	Updated_Base_	36.0	0.6277480407561902	0.6257348138471587
36	Updated_Base_	37.0	0.6278854781021089	0.6258164598301763
37	Updated_Base_	38.0	0.6280310630500402	0.6259185173089484
38	Updated_Base_	39.0	0.6278404461396393	0.6254898758981058

Continued on next page

TABLE 4 – continued from previous page

Index	Dataset	Tree Size	Weighted F1	Accuracy
39	Updated_Base_	40.0	0.6266600728033788	0.6246734160679295
40	Updated_Base_	41.0	0.628144831397881	0.6260409862834748
41	Updated_Base_	42.0	0.6268962888876656	0.625102057478772
42	Updated_Base_	43.0	0.6274145917204553	0.6253469954278249
43	Updated_Base_	44.0	0.6271949032811092	0.6250816459830176
44	Updated_Base_	45.0	0.6283038122566751	0.6260001632919661
45	Updated_Base_	46.0	0.6280390724355513	0.6259185173089484
46	Updated_Base_	47.0	0.6281462459044523	0.6261634552580013
47	Updated_Base_	48.0	0.6277646848841048	0.6257960483344219
48	Updated_Base_	49.0	0.6285854883167927	0.6266737426518615
50	Updated_21D_	1.0	0.5775574585786147	0.6050619005366121
51	Updated_21D_	2.0	0.6243887644619017	0.6340896607090041
52	Updated_21D_	3.0	0.6384444697390498	0.6384417120885622
53	Updated_21D_	4.0	0.6561755407699251	0.6604273179250997
54	Updated_21D_	5.0	0.6565664626983373	0.6683004464725849
55	Updated_21D_	6.0	0.6687921591186304	0.6779341136040338
56	Updated_21D_	7.0	0.6704231760790502	0.6807368910297038
57	Updated_21D_	8.0	0.6854585848286223	0.6924550358445656
58	Updated_21D_	9.0	0.6929933576758018	0.7001873213053337
59	Updated_21D_	10.0	0.696383576602664	0.7042154335854425
60	Updated_21D_	11.0	0.7039463597842646	0.7109477331305193
61	Updated_21D_	12.0	0.7064589039519662	0.7143561358290729
62	Updated_21D_	13.0	0.7097647297286682	0.7166941310685765
63	Updated_21D_	14.0	0.7128319109923331	0.7193983183335446
64	Updated_21D_	15.0	0.7128094635493165	0.7190462106167519
65	Updated_21D_	16.0	0.7145515040553129	0.7200321122237715
66	Updated_21D_	17.0	0.7116789132886551	0.7172293347981015
67	Updated_21D_	18.0	0.7117538239911917	0.7166518781425614
68	Updated_21D_	19.0	0.7114521425855577	0.7161589273390515
69	Updated_21D_	20.0	0.7103148961708515	0.7146659906198505
70	Updated_21D_	21.0	0.7093329533032555	0.7131167166659624
71	Updated_21D_	22.0	0.7102376006464846	0.7134829086914268
72	Updated_21D_	23.0	0.7091466702981695	0.7120603935155843
73	Updated_21D_	24.0	0.7064470218890581	0.7089618456078084

Continued on next page

TABLE 4 – continued from previous page

Index	Dataset	Tree Size	Weighted F1	Accuracy
74	Updated_21D_	25.0	0.7054794609605245	0.7077646793707131
75	Updated_21D_	26.0	0.7052632764651658	0.707285812875875
76	Updated_21D_	27.0	0.702840958846148	0.7045393726848918
77	Updated_21D_	28.0	0.7034104069316386	0.7049055647103561
78	Updated_21D_	29.0	0.7020200040284946	0.7033422064477965
79	Updated_21D_	30.0	0.7000663937301365	0.7011591386036816
80	Updated_21D_	31.0	0.7000318899903663	0.7010464641343079
81	Updated_21D_	32.0	0.6983266025987823	0.6992014196983141
82	Updated_21D_	33.0	0.6983873036553369	0.699159166772299
83	Updated_21D_	34.0	0.6968354817365744	0.6974549654230222
84	Updated_21D_	35.0	0.696134615490223	0.6967225813720933
85	Updated_21D_	36.0	0.6953065863781187	0.6958352699257757
86	Updated_21D_	37.0	0.6950185545211114	0.6954972465176547
87	Updated_21D_	38.0	0.694222301303103	0.6946521879973522
88	Updated_21D_	39.0	0.6944239733963955	0.6948352840100843
89	Updated_21D_	40.0	0.6941221959756513	0.6944972606019634
90	Updated_21D_	41.0	0.6928841167521189	0.6932296728215096
91	Updated_21D_	42.0	0.692539469402266	0.6928775651047169
92	Updated_21D_	43.0	0.6928109794180883	0.693159251278151
93	Updated_21D_	44.0	0.6924214853973947	0.6927508063266714
94	Updated_21D_	45.0	0.6925550143273675	0.6928916494133885
95	Updated_21D_	46.0	0.6930964951667452	0.6934409374515852
96	Updated_21D_	47.0	0.6930113104548286	0.693356431599555
97	Updated_21D_	48.0	0.6922943049200346	0.692624047548626
98	Updated_21D_	49.0	0.6926442152803272	0.6929761552654188
99	Updated_21D_	1.0	0.5767068441525315	0.6039351558428755
100	Updated_21D_	2.0	0.6235618823559342	0.6330051689412826
101	Updated_21D_	3.0	0.6393476590136425	0.6394416980042534
102	Updated_21D_	4.0	0.652919947045216	0.6575963718820862
103	Updated_21D_	5.0	0.6598740082623948	0.666455402036591
104	Updated_21D_	6.0	0.6713318650729277	0.6783707271728567
105	Updated_21D_	7.0	0.6735403155875928	0.6820044788101576
106	Updated_21D_	8.0	0.6761221368100988	0.6870325770059577
107	Updated_21D_	9.0	0.6887404548373453	0.6957930169997606

Continued on next page

TABLE 4 – continued from previous page

Index	Dataset	Tree Size	Weighted F1	Accuracy
108	Updated_21D_	10.0	0.6921443992530497	0.7023281362234335
109	Updated_21D_	11.0	0.7015586779439327	0.7094829650286616
110	Updated_21D_	12.0	0.7061649288884131	0.7134406557654117
111	Updated_21D_	13.0	0.7094496076263123	0.7162716018084252
112	Updated_21D_	14.0	0.7115542409572522	0.7193701497162012
113	Updated_21D_	15.0	0.7137914005512329	0.7203701356318926
114	Updated_21D_	16.0	0.7132708888918565	0.7201025337671301
115	Updated_21D_	17.0	0.7116979233834205	0.7177082012929396
116	Updated_21D_	18.0	0.7138293064613255	0.7192152223208124
117	Updated_21D_	19.0	0.7117286066731546	0.7164406135124858
118	Updated_21D_	20.0	0.7126790721564074	0.7169194800073239
119	Updated_21D_	21.0	0.7088784047337561	0.7125674286277658
120	Updated_21D_	22.0	0.7092880764093938	0.7126378501711244
121	Updated_21D_	23.0	0.7072830903018773	0.7100745059928734
122	Updated_21D_	24.0	0.7062721452195668	0.7087646652864045
123	Updated_21D_	25.0	0.7047914582440118	0.7068351149983803
124	Updated_21D_	26.0	0.7047149992784474	0.706609766059633
125	Updated_21D_	27.0	0.7034443149983823	0.705060492105745
126	Updated_21D_	28.0	0.7017767314947158	0.7031309418177208
127	Updated_21D_	29.0	0.7005965648755655	0.7018070168025803
128	Updated_21D_	30.0	0.7002660412233809	0.7013281503077421
129	Updated_21D_	31.0	0.6987495692023368	0.6996662018844805
130	Updated_21D_	32.0	0.6981979077574422	0.698961986450895
131	Updated_21D_	33.0	0.6974600787005589	0.698145096547936
132	Updated_21D_	34.0	0.6978148982021336	0.6984408670300418
133	Updated_21D_	35.0	0.6964567341855771	0.6970042675455276
134	Updated_21D_	36.0	0.6961574270459595	0.6966662441374065
135	Updated_21D_	37.0	0.6957869333110794	0.6962296305685836
136	Updated_21D_	38.0	0.694967788913904	0.695384572048281
137	Updated_21D_	39.0	0.6944618764505791	0.6948493683187561
138	Updated_21D_	40.0	0.6947840779742462	0.6951733074182054
139	Updated_21D_	41.0	0.6946432249886252	0.6950042957141449
140	Updated_21D_	42.0	0.6941059632026851	0.6944409233672765
141	Updated_21D_	43.0	0.6936914909884965	0.6940183941071253

Continued on next page

TABLE 4 – continued from previous page

Index	Dataset	Tree Size	Weighted F1	Accuracy
142	Updated_21D_	44.0	0.6939265970091343	0.6942437430458726
143	Updated_21D_	45.0	0.6934419800007067	0.693736707933691
144	Updated_21D_	46.0	0.6936095449815274	0.6939057196377516
145	Updated_21D_	47.0	0.6933205550390314	0.6936240334643174
146	Updated_21D_	48.0	0.6937861888649955	0.6940747313418121
147	Updated_21D_	49.0	0.693409520481219	0.6936944550076759

TABLE 5: GridSearch CV with our Decision Trees

SL.No	Dataset	F1 Score CV	Precision Score	Recall Score
1	Updated_Base_GS_Fin	0.6874438362483438	0.6975663973571417	0.733874918354017
2	21D_SMOTE_CV	0.7152737209142526	0.7428495995099996	0.7215250489429726
3	Base	0.9845164204741484	0.9849764321635543	0.9846606043905576
4	21D	0.9845164204741484	0.9849764321635543	0.9846606043905576

4.1. Differences in this project

The project's primary direction towards data collection was selecting features with minimal missing values because of the ordinal nature of the policy measure and hence, the weighted measures (weighted indexes) were taken into account as they had no missing values. Moreover, the dataset used in this project is the compact version proposed in [1], whereas [4] uses the country-specific dataset given for the USA, with a different time period of data collection. On the other hand, dropping values had lead to improper distribution of values across countries, which inversely affects the summary statistical measures such as mean and mode, that were used to create our models for regression tasks. [2] was the only paper that discussed their imputation strategy using moving averages for only SI indexes but it only focuses on data till May, 2020. The same group of authors published [3], focusing on Stringency Index and its relationship with daily peak cases. Similarly, [7] records its measures for multiple time periods.

[6] helps us append features to [1] from [8]. Using these features, the project focuses on supervised regression and classification models for the entire period for which [**?**] had recorded the policies, i.e., Jan 01, 2020 to Dec 31, 2022. Finally, highlights the data collection process which was not covered in the literature. While it does improve certain measures (r2: 0.977, second best: 0.691, with the majority of measures around 0.69 for daily mortality rates, and weighted F1 score of 0.7152737209142526 for classifying effective policies) the basis of comparison isn't correct as different authors have used a different set of countries, which were not clearly reported in their literature or a repository that could be accessed, and thus, have reported measures on different datasets. This reasoning stems from traditional machine learning algorithms such as linear or logistic regression that run gradient descent to improve the parameters by going through the entries in the dataset, one at a time.

The link to the repository that holds all the notebooks, comparison tables, plots, decision trees, confusion matrix is https://github.com/jeet1912/aoxcgrt.

5. Future scope of work

Finally, the project would like to state that, even though putting the models into practice gave insightful information on the efficacy of different policies, there is still room for improvement in the suggested algorithms that have already been put out, a few of which are:

- To determine whether culturally similar nations adopted similar approaches to policy in an attempt to minimize the issue, gather information on local demographics for each country and incorporate it into the analysis.
- As the WHO implements its new DPM policy, try to use the WHO dataset to understand how it might impact the existing associations pertaining to the success of the policy.
- Change the emphasis of daily trends in order to obtain a more thorough and in-depth understanding, which necessitates the selection of new characteristics.
- Utilize methods such as LSTM to record associations of higher dimensions.
- Assign additional weights to the data and determine how that affects the data further by utilizing life expectancy and demographic statistics.
- Following the research in [9] try and minimize the possible misinformation of importance of features, and checking for incorrectly assigned importance to the utilized features.

References

[1] T. Hale, N. Angrist, R. Goldszmidt, *et al.*, "A global panel database of pandemic policies (oxford covid-19 government response tracker)," *Nature Human Behaviour*, vol. 5, pp. 529–538, 2021.

[2] D. Zhu, S. R. Mishra, Y. Ma, and X. Han, "Shorter time to a high response level is associated with sooner arrival of peak daily incidence: An analysis of governments' stringency index from 151 countries." https://ssrn.com/abstract=3743993, Dec 2020.

[3] Y. Ma, S. Mishra, X. Han, and D. Zhu, "The relationship between time to a high covid-19 response level and timing of peak daily incidence: an analysis of governments' stringency index from 148 countries," *Infect Dis Poverty*, vol. 10, no. 1, p. 96, 2021.

[4] J. Stokes, A. Turner, L. Anselmi, M. Morciano, and T. Hone, "The relative effects of non-pharmaceutical interventions on wave one covid-19 mortality: natural experiment in 130 countries," *BMC Public Health*, vol. 22, no. 1, p. 1113, 2022.

[5] Y. Mahajan, S. R. Islam, M. R. Amin, and S. K. Karmaker Santu, "Data-driven estimation of effectiveness of covid-19 non-pharmaceutical intervention policies," in *2022 IEEE International Conference on Big Data (Big Data)*, (Osaka, Japan), pp. 5312–5321, 2022.

[6] S. Bouktif and N. K, "Exploring epidemiological determinants of covid-19: Clustering and correlation of multifaceted factors," in *2023 IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, (Chongqing, China), pp. 1962–1966, 2023.

[7] X. Tang, H. C. Wu, and Y. Chen, "Prediction of covid-19 confirmed cases combining the lstm model and evolutionary strategy," in *2022 4th International Academic Exchange Conference on Science and Technology Innovation (IAECST)*, (Guangzhou, China), pp. 637–641, 2022.

[8] J. Hasell, E. Mathieu, D. Beltekian, *et al.*, "A cross-country database of covid-19 testing," *Scientific Data*, vol. 7, p. 345, 2020.

[9] X. Huang and J. Marques-Silva, "The inadequacy of shapley values for explainability," 2023.