Evaluating System for Effectiveness of Mask Mandates and the Most Influential Factor in Each State in the U.S.A

Chuang Zhou1*

1School of Information Science and Engineering, Lanzhou University, Lanzhou, China
*chzhou17@lzu.edu.cn

Abstract. Over the past year, the COVID-19 outbreak deeply and thoroughly changed the way the world is, and plenty of people died because of this virus. To control the epidemic, all the state governments in the United States took the mask mandates to ask people to wear masks in public. However, many people doubt if the mask can help them prevent the spread of the virus, and there are few researches about the effectiveness of the mask mandates in each state. To tackle this issue, this paper proposed an evaluating system for the effectiveness of mask mandate in each state in the U.S.A, which can be utilized to analyze whether mask mandates can slow down the spread of the virus. And if the effectiveness of mask mandates is different in different states, this paper will continue to find out which factor can most affect the effectiveness. The experimental results demonstrated that even though the new cases in some states still increased after the mask mandates, it can be seemed that the mask mandates do control the spread of the virus in general. According to the research, the paper can find that the most influential factor to the effectiveness is the infection rate. There are also some secondary factors to the effectiveness of mask mandates. Through these researches, they can analyze the result to develop more effective mandates to control the epidemic and help governments and companies develop new products to fit the needs during the epidemic.

1. Introduction

Since the outbreak of the COVID-19 in December 2019, it has brought lots of trouble to people's normal life, factories, manufacture, and tertiary industry in all countries in the world. People are asked to halt their works and isolate themselves at home to get through the toughest days. To protect the health of citizens and recover the local economy, the United States local government enacted many policies [1], including mask mandates. However, the mask mandates met with considerable resistance in the beginning because many people think the mask is useless. Moreover, because of the severe situation of the COVID-19, there has been lots of researches in this area over this period. Nevertheless, most researches concentrate on the vaccine development and prediction model [2,3], such as The Global COVID-19 Prediction System, established by a team of the Collaborative Innovation Center of Western Ecological Safety (CIWES) led by Professor Huang, the director of the center. But there are also some researches about the effect of mask in preventing COVID-19 infection. For instance, in those work of Wang et al. [4] and Daniel P et al. [5], they do some research to illustrate that the mask, especially some kinds of masks, has a good effect preventing COVID-19 infection. However, their researches focus on the function of internal properties of masks like types of mask in preventing the spread of the virus. And few researches use machine learning to analyze the external factors that influence the effectiveness of the mask.
This paper focuses on evaluating the effectiveness of mask mandates in different states and find the most influential external factor that influences the effectiveness of the mask mandates. Based on the data recording the new cases per day before mask mandates and after mask mandates from some of the states in the United States, this research uses the LES algorithm [6] to calculate the increase cases per day during the two different periods. By comparing those two results from the LES algorithm, this research can figure out a parameter that describes the mask mandates’ effectiveness in this state. Through those parameters, the effectiveness of mask mandates in each state can be evaluated. Finally, importing some kinds of external factors in each state, the research uses the ID3 algorithm to analyze the most influential factor to the effectiveness of mask mandates.

The main contributions of this work can be summarized as follows:
1. This paper develops a kind of evaluating system for the effectiveness of mask mandates in each state in the U.S.A.
2. The external factors and the effectiveness of the mask itself are linked in this work.
3. This paper finds the most influential factor to the different effects of mask mandates in different states.
4. Practical advice is provided for governments and companies during the epidemic [7].

The rest of this paper is organized as follows. In Sect. 2, some necessary methodologies are discussed. In Sect. 3, the effectiveness in a different state is analyzed and then the most influential factor to the effectiveness of mask mandates is found. Finally, Sect. 4 concludes this paper.

2. Methodology

2.1. LES algorithm

Assuming that the independent variable is \( \overrightarrow{x}_i = [x^{i_1}, \ldots, x^{i_n}] \) and the model's parameters are \( \overrightarrow{\beta} = [\beta_0, \ldots, \beta_m] \), then the model's prediction would be \( y_i \approx \beta_0 + \sum_{j=1}^{m} \beta_j \times x^j_i \). If \( \overrightarrow{x}_i \) is extended to \( \overrightarrow{x}_i = [1, x^{i_1}, \ldots, x^{i_n}] \) then \( y_i \) would become a dot product of the parameter and the independent variable, i.e. \( y_i \approx \sum_{j=1}^{m} \beta_j \times x^j_i = \overrightarrow{\beta} \cdot \overrightarrow{x}_i \).

In the least-squares setting, the optimum parameter is defined as such that minimizes the sum of mean squared loss:

\[
\overrightarrow{\beta} = \arg \min_{\beta} L(D, \overrightarrow{\beta}) = \arg \min_{\beta} \sum_{i=1}^{n} (\overrightarrow{\beta}, X_i - y_i)^2
\]  

(1)

Now putting the independent and dependent variables in matrices \( X \) and \( Y \), respectively, the loss function can be rewritten as:

\[
L(D, \overrightarrow{\beta}) = \|X \overrightarrow{\beta} - Y\|^2 = (X \overrightarrow{\beta} - Y)^T (X \overrightarrow{\beta} - Y) = Y^T Y - Y^T X \overrightarrow{\beta} - \overrightarrow{\beta}^T X^T Y + \overrightarrow{\beta}^T X^T X \overrightarrow{\beta}
\]

(2)

As the loss is convex, the optimum solution lies at gradient zero. The gradient of the loss function is (using Denominator layout convention):

\[
\frac{\partial L(D, \overrightarrow{\beta})}{\partial \overrightarrow{\beta}} = -2X^T Y + 2X^T X \overrightarrow{\beta}
\]

(3)

Setting the gradient to zero produces the optimum parameter:

\[
\frac{\partial L(D, \overrightarrow{\beta})}{\partial \overrightarrow{\beta}} = 0 \Rightarrow -2X^T Y + 2X^T X \overrightarrow{\beta} = 0 \Rightarrow \overrightarrow{\beta} = (X^T X)^{-1} X^T Y
\]

(4)
2.2. Decision tree
A decision tree builds classification models in the form of a tree structure. The result is a tree with decision nodes and leaf nodes.

2.2.1. Terminology.

1. Root Node: It represents the entire population or sample, and this further gets divided into two or more homogeneous sets.
2. Splitting: It is a process of dividing a node into two or more sub-nodes.
3. Decision Node: When a sub-node splits into further sub-nodes, it is called a decision node.
4. Leaf/Terminal Node: Nodes with no children (no further split) is called Leaf or Terminal node.
5. Branch/Sub-Tree: A subsection of the decision tree is called a branch or sub-tree.
6. Parent and Child Node: A node, which is divided into sub-nodes, is called the parent node of sub-nodes, whereas sub-nodes are the child of the parent node.

2.2.2. ID3 algorithm formulation. This algorithm employs a top-down, greedy search through the space of possible branches with no backtracking. A decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values. ID3 algorithm uses entropy to calculate the homogeneity of a sample. If the sample is completely homogeneous, the entropy is zero, and if the sample is equally divided, then it has an entropy of one.

Entropy using the frequency table of one attribute:

$$E(s) = \sum_{i=1}^{c} -p_i \log_2 p_i$$  \hspace{1cm} (5)$$

Entropy using the frequency table of over one attributes:

$$E(T, X) = \sum_{c \in X} P(c)E(c)$$  \hspace{1cm} (6)$$

Where $P(c)$ indicates the probability of parameter $c$, and $E(c)$ indicates the entropy of the parameter $c$.

Fig. 2 shows the distribution function of the entropy, which presents the results of entropy in various attributes with different frequencies.
2.2.3. **Information gain.** The information gain is based on the decrease in entropy after a data-set is split on an attribute. Constructing a decision tree is all about finding an attribute that returns the highest information gain (i.e., the most homogeneous branches).

\[
Gain(T, X) = Entropy(T) - Entropy(T, X)
\]

2.2.4. **Building of decision tree.**
Step 1: Calculate the entropy of the target.
Step 2: The dataset is then split into different attributes. The entropy for each branch is calculated. Then it is added proportionally to get total entropy for the split. The resulting entropy is subtracted from the entropy before the split. The result is the Information Gain or decrease in entropy.
Step 3: Choose the attribute with the largest information gain as the decision node, divide the dataset by its branches and repeat the same process on every branch.
Step 4a: A branch with an entropy of 0 is a leaf node.
Step 4b: A branch with an entropy of more than 0 needs further splitting.
Step 5: The algorithm is run recursively on the non-leaf branches until all data is classified.

3. **Experiments**

3.1. **Dataset and pretreatment**

3.1.1. **New cases per each day in different states during different period.**

| Date     | Period | State   | Cases | Death |
|----------|--------|---------|-------|-------|
| 2020/5/16| pre_mask| Virginia| 1010  | 25    |
| 2020/5/30| incubation| Virginia| 1078  | 12    |
| 2020/6/17| post_mask| Virginia| 442   | 13    |

This dataset in Tab. 1 includes the new cases and deaths in a particular date and state. And according to the time of the mandates' enactment, dates can be classified into three periods- the pre_mask period, the incubation period, and the post_mask period. The pre_mask period means the time before the enactment.
of mandates, and the post_mask period means the time after the enactment and the incubation period. Both the pre_mask period and post_mask period has 14 days in each state. The incubation period refers to the time the infection lives in the body before it shows symptoms. On average, this period has 5 days. Later researches will ignore this period to get more reasonable results.

3.1.2. Temperature.

Tab. 2 The dataset about the temperature in different states

| State   | Temperature | Rank |
|---------|-------------|------|
| Alaska  | -3          | 50   |
| Arizona | 15.7        | 10   |
| Arkansas| 15.8        | 9    |

The temperature dataset in Tab. 2 is based on the average temperature of each state in the United States, and to get relative distinction in different states' temperature, the research divided all states into three categories by their temperature ranks. Temperature classification standard:

(1) Rank <=16: Hot
(2) 16<Rank<=34: Warm
(3) Rank>34: Cold

3.1.3. Staffed bed.

Tab. 3 The dataset about the number of hospitals and staffed beds in different states

| State   | Hospitals | Staffed beds |
|---------|-----------|--------------|
| Alaska  | 10        | 1274         |
| Arizona | 77        | 13789        |
| Arkansas| 51        | 7999         |

This dataset in Tab. 3 is based on total hospitals and staffed beds in the United States. The number of hospitals and the number of the staffed beds is a positive correlation, and the number of staffed beds can better show the medical situation during the epidemic. As a result, this research decides to use the state's total cases in a month to divide the summation of staffed beds in the state, then get cases per bed of each state. Then use all cases per bed in different states to classify each state.

Cases per bed classification standard:

(1) Cases per bed<=1.5: Vacant
(2) 1.5<Cases per bed<=3.5: Normal
(3) Cases per bed >3.5: Redundant

Cases per bed in each state in Fig. 3:
3.1.4. **Infection rate.**

**Tab. 4 The dataset about the population in different states**

| State   | Population   | Proportion |
|---------|--------------|------------|
| California | 39,512,223   | 11.91%     |
| Texas    | 28,995,881   | 8.74%      |
| Florida  | 21,477,737   | 6.47%      |

This dataset in Tab. 4 is based on the population of each state in the United States. This research decides to use the state's total cases in a month to divide the state's population and get the infection rate. Then this paper uses all the infection rates in different states to do a MIN-MAX Normalization. Finally, this paper uses the updated infection rates to classify each state by a uniform classification standard.

MIN-MAX Normalization can be formulated as:

$$x_{\text{normalization}} = \frac{x - x_{\text{Min}}}{x_{\text{Max}} - x_{\text{Min}}} \quad (8)$$

**Infection rate classification standard:**

1. Infection rate < = 0.2: Low
2. 0.2 < Infection rate < = 0.45: Medium
3. Infection rate > 0.45: High

Infection rate in each state in Fig. 4:
3.1.5. Population density.

Tab. 5: The dataset about the population density in different states

| State    | Density |
|----------|---------|
| New Jersey | 1218    |
| Puerto Rico | 1046   |
| Rhode Island | 1021  |
| ......     | ......  |

The Population density dataset in Tab. 5 is based on the average population density in the United States, and to get relative distinction in different states' population density, the research divided all states into three categories by their number of residents per square mile in a uniform classification standard.

Population density classification standard:

(1) Number <=100: Low Density
(2) 100< Number <=250: Normal Density
(3) Number >250: High Density

3.2. Experimental results visualization and analysis

3.2.1. Efficiency parameter. According to the dataset from Tab.1, the research got the new cases per day in different states during three different periods, pre_mask period, incubation period, and post_mask period. The research uses the LES algorithm in both the pre_mask period and post_mask period to get two linear functions in one particular state. Fig. 5 and Fig. 6 beneath are the result of Virginia. The y-axis indicates the new cases on one particular day, and the x-axis represents the number of days from the start date.
After comparing two linear functions of two periods in Virginia from Fig. 5 and Fig. 6, it is easy to find that the slope of the LES function decreases after the mask mandates, which means that the mask mandates do help Virginia to control the epidemic.

The research continues to do the same thing to every state. As a result, the research yields two functions in each state. Then the research defines the slope of the LES function before mandating mask policy as $K_1$ and the slope of the LES function after mandating mask policy as $K_2$.

And then, the research uses $K_1$ and $K_2$ to calculate the efficiency parameter:

$$\text{Efficiency parameter } = K_2 - K_1$$  \hspace{1cm} (9)

The lower the efficiency parameter is, the better effect of the mask mandates in the state.

3.2.2. Three kinds of states for their effectiveness of the mask mandates. The research uses all of the efficiency parameters in the figure to calculate the average number of the efficiency parameter: $-21.752030578117534$. And Fig. 7 shows all the efficiency parameters in different states.
Fig. 7 Efficiency parameter in each state

According to Fig. 7 and the average efficiency parameter, the research can come to a conclusion: According to the average of Efficiency parameters, mandates play positive roles in controlling the COVID-19. And Mandates do help some states to control the COVID-19 but are also less helpful in some states.

Using the efficiency parameter, the research sets up a uniform standard to classify each state into good, normal, and bad three categories, representing the effect of mask mandate in each state. Fig. 8 shows the proportion of three categories in all states.

Classification standard of the effect of mask mandate:

(1) Efficiency parameters <=-10: good
(2) -10<Efficiency parameters<10: normal
(3) Efficiency parameters >=10: bad

3.2.3. Four attributes of each state about effectiveness of the mask mandates. To find out why the mask mandates have different effects in different states [8], the research chooses four relevant external factors - temperature, staffed bed, infection rate, and population density as attributes of the effect of mask mandates.
The reasons for choosing these four factors can be summarized as follows:

(1) Temperature: Viruses can survive longer in cold temperatures, which increases the chance that people are exposed to them [9].

(2) Staffed bed: More staffed beds can contain more patients in the hospitals instead of isolating at home.

(3) Infection rate: The higher the infection rate, the greater the chance that people will contact a sick person [10].

(4) Population density: High density means that it is more difficult for people to keep social distancing.

All the states get four attributes from the datasets from 3.1.2-3.1.5, which have already been pretreated.

3.2.4. Result of decision tree. The research has already gotten the label about the effectiveness of mask mandates of each state and four attributes in all states. Therefore, the research uses the ID3 algorithm to draw the final result of the decision tree. All the building processes are mentioned in 2.3.

The research calculates each attribute's information gain, which is presented in Fig. 9. The infection rate has the highest information gain, which means infection rate is the most influential factor in the effectiveness of the mask mandates.

![Fig. 9 Information gain of each attribute](image)

The research continues to use the ID3 algorithm until the final decision tree which is exhibited in Fig. 10 is obtained. Through this decision tree, the research can find out that infection rate is the most influential external factor to the effect of mask mandate. The higher the infection rate is, the better effect the mask mandates have. Besides, it is clear to see the influence of temperature, population density, and cases per bed on the effect of mask mandate. It is worth noting that in the state with high and low infection rates, the colder the temperature is, the worse the effect of mask mandates is because the cold temperature keeps the viruses surviving longer. But in the states with a medium infection rate, the function of the temperature reverses. Its possible reason is that when the infection rate is medium, which means that the most influential external factor has less function than the states having extreme infection rates, the function of temperature does not relate to the surviving time of the virus. If the temperature is too hot, the number of people who follow the mask mandates will decrease, which leads to this interesting situation. And as for the attribute-cases per bed, the factor has the function as the research expecting. If the number of cases per bed is too high, it means that the hospitals are redundant. Numerous infected people cannot get treat in the hospital, they are asked to stay at home, which increases the possibility of infection in public. As a result, the mask has the same function as the condition with a high infection rate. The population density has different functions when the precondition is different, and it needs further research to know its function in detail. In a word, according to the final result of the
decision tree in Fig. 10, the research can help the government to come up with more reasonable solutions to control the COVID-19.

![Decision Tree Illustration](image)

**Fig. 10 Illustration of decision tree**

4. **Conclusion**

This paper proposed a reasonable evaluating system for the effectiveness of mask mandates in each state in the U.S.A and finds out the most influential external factor of the mask to the effectiveness of mask mandates. Through the LES algorithm, the research illustrates the mask mandate does slow the spread of the virus and classify all the state with the data about the new cases before and after the effective day of the mandate, which can help the government take multilayer policies to find the balance between controlling the epidemic and recovering the economy. Furthermore, this research also uses the ID3 algorithm to analyze the external factors' relationship with the effectiveness of mask mandates. The research results show that the infection rate is the most influential external factor to the effectiveness of mask mandates, and some secondary factors also affect the effectiveness. By analyzing those relevant factors, governments can set more reasonable policies to control the epidemic, and some companies can produce new products to satisfy the needs during the epidemic.

In the future, the government can use this evaluating system to assess each state's situation of the epidemic and take appropriate measures by its own characters. For example, those states, which are in high infection rate should take more strict mandates to require citizens to wear a mask, because based on the current work and problem, the higher the infection rate the state is, the better effect the mask mandates are. Besides this, each state, especially the state with low temperature, need to disinfect public place with the crowd. From the analysis of experimental results, it is clear that the mask can only prevent the spread of the virus in the atmosphere. If people touch the remain of the virus by hands or other parts of the body, it can hardly have a good effect in preventing. To better control the epidemic, it is necessary to disinfect the public place, even though people have already worn the masks. Furthermore, by comparing and analyzing the experimental results, it is a good chance for some materials industries to produce a kind of heat tolerance mask on account of the new need during the epidemic. If this kind of mask can be manufactured, it can improve the wear rate of the mask in those hot states and take advantage of the potential markets.
References

[1] A. Adekunle et al, "Delaying the COVID-19 epidemic in Australia: evaluating the effectiveness of international travel bans" Australian and New Zealand Journal of Public Health, vol. 44, no. 4, pp. 257-259, 2020.

[2] S. Nagpal, V. A. Athavale, A. K. Saini and R. Sharma, "Indian Health Care System is Ready to Fight Against COVID-19 A Machine Learning Tool for Forecast the Number of Beds," 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), pp. 61-65, 2020.

[3] R. Y. Wang, T. Q. Guo, L. G. Li, J. Y. Jiao and L. Y. Wang, "Predictions of COVID-19 Infection Severity Based on Co-associations between the SNPs of Co-morbid Diseases and COVID-19 through Machine Learning of Genetic Data," 2020 IEEE 8th International Conference on Computer Science and Network Technology (ICCSNT), pp. 92-96, 2020.

[4] Y. Wang, Z. Deng and D. Shi, "How effective is a mask in preventing COVID-19 infection?" MEDICAL DEVICES & SENSORS, vol. 4, no. 1, 2021.

[5] R. Daniel P, H. Karisa K, V. H. Paul, F. Brett, O. Patrick and M. Nicholas M, "The "double eights mask brace" improves the fit and protection of a basic surgical mask amidst COVID-19 pandemic" Journal of the American College of Emergency Physicians Open, vol. 2, no. 1, 2021.

[6] A. Battistini, A. Cammi, S. Lorenzi, M. Colombo and M. Fairweather, "Development of a CFD – LES model for the dynamic analysis of the DYNASTY natural circulation loop" Chemical Engineering Science, vol. 237, 2021.

[7] X. Zhang and W. Mildred, "COVID-19 Policy Differences across US States: Shutdowns, Reopening, and Mask Mandates" International Journal of Environmental Research and Public Health, vol. 17, no. 24, pp. 9520-9520, 2020.

[8] D. Sharoda et al, "Differences in rapid increases in county-level COVID-19 incidence by implementation of statewide closures and mask mandates — United States, June 1–September 30, 2020" Annals of Epidemiology, vol. 57, pp. 46-53, 2021.

[9] S. Ihsan et al, "Risk factors affecting icu admission in COVID 19 patients; could air temperature be an effective factor?" International journal of clinical practice, pp. e13803-e13803, 2020.

[10] Credit Kevin, "Neighbourhood inequity: Exploring the factors underlying racial and ethnic disparities in COVID-19 testing and infection rates using ZIP code data in Chicago and New York" Regional Science Policy & Practice, vol. 12, no. 6, pp. 1249-1271, 2020.