US mask policy during COVID-19 evaluation

1. Reporter

Sam(Chuang Zhou) & Jim(Jin Cui)

Group 4

US mask policy during COVID-19 evaluation

2. Executive Summary

Over the past year the COVID-19 outbreak deeply and thoroughly changed the way the world is. Through the way of machine learning, now we are able to find and to probe into the data about corona-virus spreading patterns, suiting the remedy to the case, to launch targeted prevention policies and minimize the economic loss under the premise of control the spread of the virus on a large scale. Therefore, we choose to analyze the spread cases of COVID-19 in the United States, and expect to find out if the mask mandates help the government to control the problems and which factor most affect the effect of the mandate. All the data are from the statistics of New York Times cited by Google.

3. Introduction

Our goal is to find how state mask policies and various disease control policies influenced epidemic outbreak. We use LES, K-means and ID3 algorithm to fit the model we build in a special way in the project. All data we collected from the internet are not intended for commercial use therefore there is no copyright issue. The tools we used are all free. We honestly believe that our project have found the clues about virus propagation pattern. Precisely because of powerful machine learning tools we can try to understand viruses and minimize its damage.

4. Methods

4.1 LES algorithm

Assuming that the independent variable is $\vec{x_i} = [x_1^i, x_2^i, \ldots, x_m^i]$ and the model's parameters are $\vec{\beta} = [\beta_0, \beta_1, \ldots, \beta_m]$, then the model's prediction would be $y_i \approx \beta_0 + \sum_{j=1}^{m} \beta_j \times x_j^i$. If Xi is extended to $\vec{x_i} = [1, x_1^i, x_2^i, \ldots, x_m^i]$ then Yi would become a dot product of the parameter and the independent variable, i.e. $y_i \approx \sum_{j=0}^{m} \beta_j \times x_j^i = \vec{\beta} \cdot \vec{x_i}$. In the least-squares setting, the optimum parameter is defined as such that minimizes the sum of mean squared loss: $\hat{\vec{\beta}} = \arg\min_{\vec{\beta}} L\left(D, \vec{\beta}\right) = \arg\min_{\vec{\beta}} \sum_{i=1}^{n} \left(\vec{\beta} \cdot \vec{x_i} - y_i\right)^2$

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

$$
\begin{aligned}
L\left(D, \vec{\beta}\right) &= \|X\vec{\beta} - Y\|^2 \\
&= \left(X\vec{\beta} - Y\right)^{\mathsf{T}} \left(X\vec{\beta} - Y\right) \\
&= Y^{\mathsf{T}}Y - Y^{\mathsf{T}}X\vec{\beta} - \vec{\beta}^{\mathsf{T}}X^{\mathsf{T}}Y + \vec{\beta}^{\mathsf{T}}X^{\mathsf{T}}X\vec{\beta}
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{\partial L\left(D, \vec{\beta}\right)}{\partial \vec{\beta}} &= \frac{\partial \left(Y^{\mathsf{T}}Y - Y^{\mathsf{T}}X\vec{\beta} - \vec{\beta}^{\mathsf{T}}X^{\mathsf{T}}Y + \vec{\beta}^{\mathsf{T}}X^{\mathsf{T}}X\vec{\beta}\right)}{\partial \vec{\beta}} \\
&= -2X^{\mathsf{T}}Y + 2X^{\mathsf{T}}X\vec{\beta}
\end{aligned}
$$

Setting the gradient to zero produces the optimum parameter:

$$
\begin{aligned}
-2X^{\mathsf{T}}Y + 2X^{\mathsf{T}}X\vec{\beta} &= 0 \\
\Rightarrow X^{\mathsf{T}}Y &= X^{\mathsf{T}}X\vec{\beta} \\
\Rightarrow \hat{\vec{\beta}} &= \left(X^{\mathsf{T}}X\right)^{-1}X^{\mathsf{T}}Y
\end{aligned}
$$

4.2 K-Means clustering

The most common algorithm uses an iterative refinement technique. Due to its ubiquity, it is often called "the *k*-means algorithm"; it is also referred to as Lloyd's algorithm, particularly in the computer science community. It is sometimes also referred to as "naïve *k*-means", because there exist much faster alternatives.

Given an initial set of *k* means *m*1(1),…,*mk*(1) (see below), the algorithm proceeds by alternating between two steps:

Assignment step: Assign each observation to the cluster with the nearest mean: that with the least squared Euclidean distance. (Mathematically, this means partitioning the observations according to the Voronoi diagram generated by the means.)

$$
S_i^{(t)} = \left\{ x_p : \left\| x_p - m_i^{(t)} \right\|^2 \le \left\| x_p - m_j^{(t)} \right\|^2 \ \forall j, 1 \le j \le k \right\},
$$

where each x_p is assigned to exactly one $S^{(t)}$ even if it could be assigned to two or more of them.

Update step: Recalculate means for observations assigned to each cluster.

$$
m_i^{(t+1)} = \frac{1}{\left|S_i^{(t)}\right|} \sum_{x_j \in S_i^{(t)}} x_j
$$

The algorithm has converged when the assignments no longer change. The algorithm is not guaranteed to find the optimum.

The algorithm is often presented as assigning objects to the nearest cluster by distance. Using a different distance function other than (squared) Euclidean distance may prevent the algorithm from converging. Various modifications of *k*-means such as spherical *k*-means and k-medoids have been proposed to allow using other distance measures.

4.3 Information Gain

$$IG(S,A) = \mathrm{H}(S) - \sum_{t \in T} p(t)\mathrm{H}(t) = \mathrm{H}(S) - \mathrm{H}(S|A).$$

Where,

- $\mathrm{H}(S)$ – Entropy of set S
- T – The subsets created from splitting set S by attribute A such that $S = \bigcup_{t \in T} t$
- $p(t)$ – The proportion of the number of elements in t to the number of elements in set S
- $\mathrm{H}(t)$ – Entropy of subset t

In ID3, information gain can be calculated (instead of entropy) for each remaining attribute. The attribute with the **largest** information gain is used to split the set S on this iteration.

4.4 Decision Tree

The ID3 algorithm begins with the original set as the root node. On each iteration of the algorithm, it iterates through every unused attribute of the set S and calculates the entropy H(s) or the information gain IG(s) of that attribute. It then selects the attribute which has the smallest entropy (or largest information gain) value. The set is then split or partitioned by the selected attribute to produce subsets of the data. (For example, a node can be split into child nodes based upon the subsets of the population whose ages are less than 50, between 50 and 100, and greater than 100.) The algorithm continues to recurse on each subset, considering only attributes never selected before.

Recursion on a subset may stop in one of these cases:

· every element in the subset belongs to the same class; in which case the node is turned into a leaf node and labelled with the class of the examples.

· there are no more attributes to be selected, but the examples still do not belong to the same class. In this case, the node is made a leaf node and labelled with the most common class of the examples in the subset.

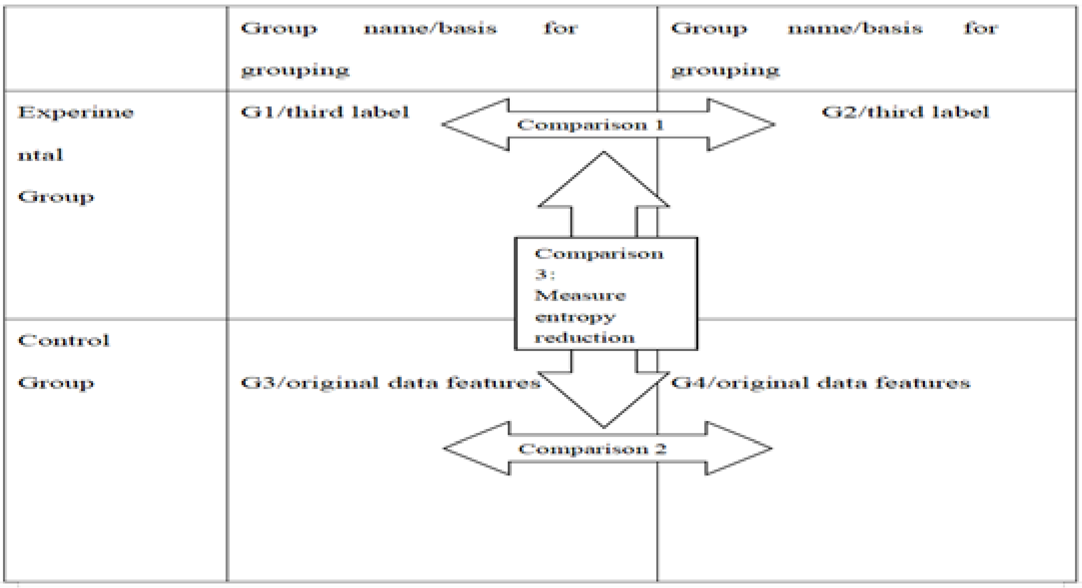
· there are no examples in the subset, which happens when no example in the parent set was found to match a specific value of the selected attribute. An example could be the absence of a person among the population with age over 100 years. Then a leaf node is created and labelled with the most common class of the examples in the parent node's set.

Throughout the algorithm, the decision tree is constructed with each non-terminal node (internal node) representing the selected attribute on which the data was split, and terminal nodes (leaf nodes) representing the class label of the final subset of this branch.

5. Result

5.1 Determine Effectiveness of Mask Mandates in each state

Data structure：



Experimental group: add a label to the label list, such as before and after the promulgation of the mask policy, divide the original data into two groups, use linear fitting to reflect the features of the data set respectively, and compare the parameters of the fitting curve, and variation factor is delta K(H(C)).

Control group: In order to exclude the influence of other factors, such as vaccination or large-scale sports events, the unsupervised classification method is adopted to classify the two-dimensional data directly. The original data is also divided into two groups this time. Linear fitting is used to reflect the features of the data set respectively, and the parameters of the fitting curve are compared, and variation factor is delta K(H(C|T)).

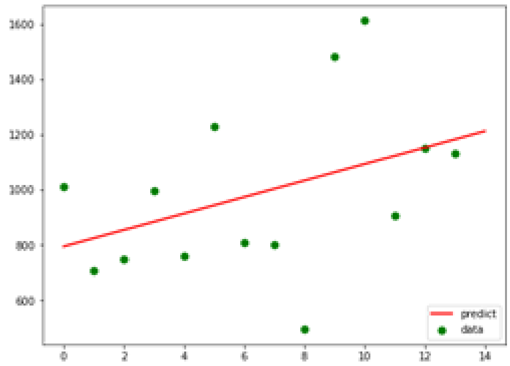
The difference of the fitting results of the experimental group and the difference of fitting results of the control group will be compared. If the difference of the experimental group is greater than or far greater than the difference of the control group, it indicates that the information contained in this label, namely entropy reduction, is larger, and the policy corresponding to this label is more effective.

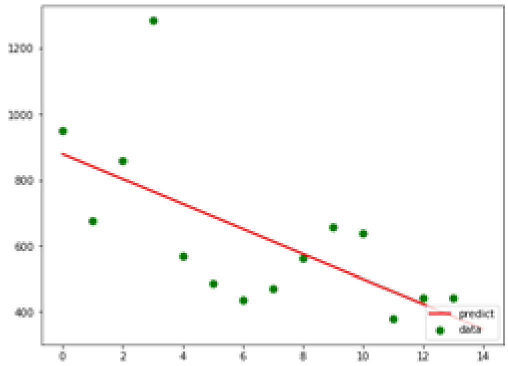
$IG(T)=H(C) - H(C|T)$

We use rate of relative change of information gain in this case.

$RC = (IG(T) - H(C)) / IG(T)$

Example State: Virginia





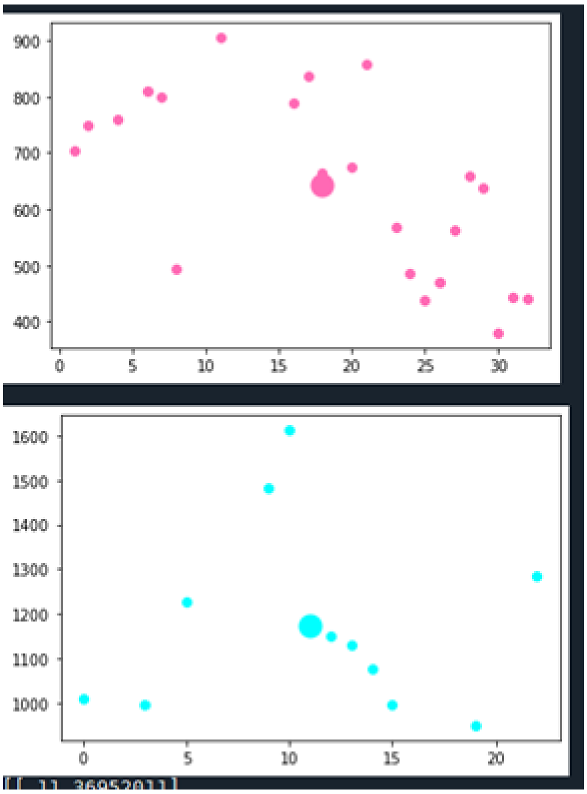
before mandating mask policy:

y1 = 29.91868x + 793.6

after mandating mask policy:

y2 = -37.95165x + 878.82857

H(C) = delta k = |k2-k1| = 67.87032967032962



Group1 LES result:

y1 = -11.13278816x + 853.98948598

Group2 LES result:

y2 = -6.11191099 + 1265.46531414

H(C|T) = delta k = |K2-K1| = 5.02087717

(67.87-5.02)/67.87 = 92.6%>>50%

The policy corresponding to this label(mask mandatory) is indeed effective.

with mask label LES result(G1 and G2 result):

state:CaliforniaK2-K1= 224.16483516483487

state:ConnecticutK2-K1= -17.21098901098897

state:DelawareK2-K1= -14.27472527472527

state:HawaiiK2-K1= 0.9186813186813195

state:IllinoisK2-K1= -145.98021978021976

state:KansasK2-K1= -17.982417582417604

state:KentuckyK2-K1= 7.107692307692307

state:MaineK2-K1= -0.6791208791208803

state:MarylandK2-K1= -1.1142857142857352

state:MassachusettsK2-K1= 104.90109890109898

state:MichiganK2-K1= 147.421978021978

state:NevadaK2-K1= -2.0527472527472845

state:New JerseyK2-K1= -167.82637362637362

state:New MexicoK2-K1= 0.5142857142857169

state:New YorkK2-K1= -296.51868131868093

state:North CarolinaK2-K1= 20.698901098901047

state:OregonK2-K1= 1.046153846153837

state:PennsylvaniaK2-K1= 5.593406593406627

state:Rhode IslandK2-K1= 6.347252747252755

state:TexasK2-K1= -287.8109890109891

state:VirginiaK2-K1= -67.87032967032962

state:WashingtonK2-K1= 2.70109890109887

state:West VirginiaK2-K1= -2.391208791208796

without mask label k-means clustering and LES result(G3 and G4 result):

state:CaliforniaK2-K1= [[-17.62382872]]

state:ConnecticutK2-K1= [[138.55352539]]

state:DelawareK2-K1= [[-1.48995755]]

state:HawaiiK2-K1= [[-0.34217362]]

state:IllinoisK2-K1= [[40.95293128]]

state:KansasK2-K1= [[-16.56702489]]

state:KentuckyK2-K1= [[-8.29173646]]

state:MaineK2-K1= [[1.75946716]]

state:MarylandK2-K1= [[-7.52532368]]

state:MassachusettsK2-K1= [[40.18064516]]

state:MichiganK2-K1= [[-8.23306452]]

state:NevadaK2-K1= [[4.26402872]]

state:New JerseyK2-K1= [[7.00590169]]

state:New MexicoK2-K1= [[-0.61546622]]

state:New YorkK2-K1= [[162.63649512]]

state:North CarolinaK2-K1= [[0.34159684]]

state:OregonK2-K1= [[3.07859755]]

state:PennsylvaniaK2-K1= [[9.50074248]]

state:Rhode IslandK2-K1= [[-1.09629328]]

state:TexasK2-K1= [[-35.08705931]]

state:VirginiaK2-K1= [[5.02087717]]

state:WashingtonK2-K1= [[-11.36952011]]

state:West VirginiaK2-K1= [[-1.77532935]]

These states results are generated by our code final-LES.ipynb and final-kmeans.ipynb, saved in the text file infection rate LES.txt and infection rate kmeans.txt.

Using EXCEL for further processing the results, and results are as follows.

states contain label LES result: average |k2-k1| = 67.0925

states without mask label k-means clustering and LES result: average |k2-k1| = 22.7526

(67.0925-22.7576)/67.0925 = 66.08% >50%

It indicates that mandating masks had a positive impact on epidemic control. On the whole, it had played its role pretty well.

The following table shows that each state's mask policy efficiency

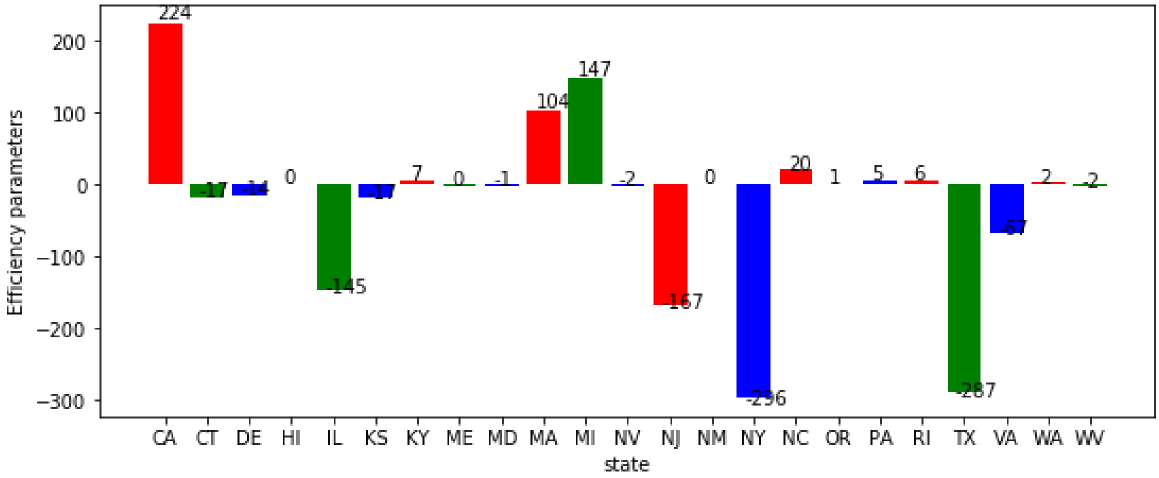
compared to 50%(TRUE is higher):

state:California	TRUE
state:Connecticut	FALSE
state:Delaware	TRUE
state:Hawaii	TRUE
state:Illinois	TRUE
state:Kansas	FALSE
state:Kentucky	FALSE
state:Maine	FALSE
state:Maryland	FALSE
state:Massachusetts	TRUE
state:Michigan	TRUE
state:Nevada	FALSE
state:New Jeresy	TRUE
state:New Mexico	FALSE
state:New York	FALSE
state:North Carolina	TRUE
state:Oregon	FALSE
state:Pennsylvania	FALSE
state:Rhode Island	TRUE
state:Texas	TRUE
state:Virginia	TRUE
state:Washington	FALSE
state:West Virginia	FALSE

5.2 Find the factor that most affect the effect of the Mask Mandates

Using LES algorithm in the data before mask mandate and after mask mandate, we can got two slop K2 and K1, which mean that the increasing rate of new cases in each day. To reduce K2 by K1, it becomes the efficiency parameter-the smaller the parameter, the better the mandate is. And the average of the efficiency parameter is -21.752030578117534 , which means that mandates play positive roles in controlling the COVID-19 in general.

efficiency parameter of each state:



Use the efficiency parameter, we set up a uniform standard to classify each state into good, normal and bad three categories, representing the effect of mask mandate in each state.

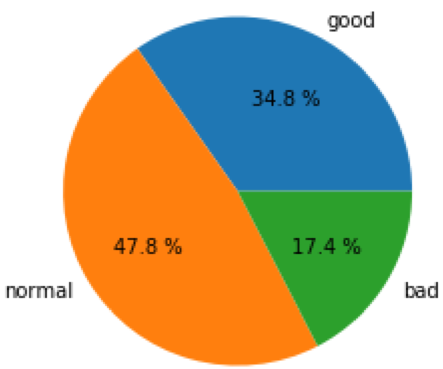
classification standard of the effect of mask mandate:

Efficiency parameters <=-10: good

-10<Efficiency parameters<10 : normal

Efficiency parameters >=10: bad

distribution of 3 categories:



To find out why the mask mandates have different effect in different state, we choose four relevant factors - temperature, staffed bed, infection rate and population density as the attribute of the effect of mask mandates. We use ID3 algorithm to draw a decision tree to find the most influential factor to the effect of mask mandate.

reasons for choosing these four factor:

Temperature: Viruses can survive longer in cold temperatures, which increases the chance that people will be exposed to them

Staffed bed: More staffed bed can contain more patients in the hospitals instead of isolating at home

Infection rate: The higher the infection rate, the greater the chance that people will come into contact with a sick person

Population density: High density means that it’s more difficult for people to keep social distancing

classification standard of four factors:

Temperature:

To get the relative temperature difference in each state, we divide the states into three categories by their temperature ranks.

Rank <=16: Hot

17<Rank<=34:   Warm

Rank>34: Cold

Staffed bed:

Use the state’s total cases in a month to divide the summation of staffed beds in the state, then get the cases per bed. Then use all the cases per bed in different state to classify each state.

Cases per bed<=1.5: Vacant

1.5<Cases per bed<=3.5:Normal

Infection>3.5:   Redundant

Infection rate:

Use the state’s total cases in a month to divide the population of the state, then get the infection rate. Then use all the infection rate in different state to do a normalization. Finally, use the updated infection rate to classify each state.

Infection rate <=0.2: Low

0.2<Infection rate<=0.45: Medium

Infection>0.45: High

Population density:

To get the relative population density difference in each state, we divide the states into three categories by their number of residents per square mile.

Number <=100:   Low Density

100< Number <=250: Normal Density

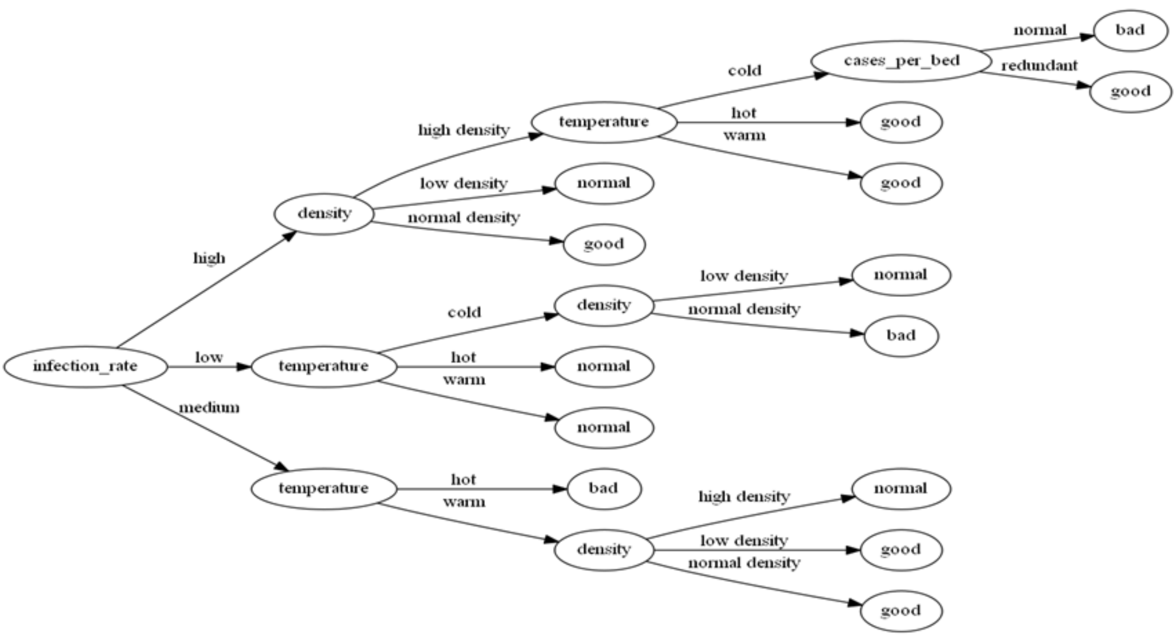
Number >250:   High Density

Use all of the attribute and the effect of mask mandates to produce a table:

state	effect_of_mask	infection_rate	cases_per_bed	density	temperature
California	bad	medium	normal	high density	hot
Connecticut	good	high	normal	high density	warm
Delaware	good	high	normal	high density	hot
Hawaii	normal	low	vacant	normal density	hot
Illinois	good	high	normal	normal density	warm
Kansas	good	medium	normal	low density	warm
Kentucky	normal	low	vacant	normal density	hot
Maine	normal	low	vacant	low density	cold
Maryland	normal	medium	normal	high density	warm
Massachusetts	bad	high	normal	high density	cold
Michigan	bad	low	vacant	normal density	cold
Nevada	normal	high	normal	low density	warm
New Jersey	good	high	redundant	high density	warm
New Mexico	normal	low	vacant	low density	warm
New York	good	high	redundant	high density	cold
North Carolina	bad	medium	normal	normal density	hot
Oregon	normal	low	vacant	low density	warm
Pennsylvania	normal	medium	vacant	high density	warm
Rhode Island	normal	medium	normal	high density	warm
Texas	good	high	redundant	normal density	hot
Virginia	good	medium	normal	normal density	warm
Washington	normal	low	redundant	high density	warm
West Virginia	normal	low	vacant	low density	warm

Use ID3 in those data and get the final decision tree.

Decision tree:



Through this decision tree, we can find out that infection rate is the most influential factor to the effect of mask mandate. Besides, we can also see the influence of temperature, population density and cases per bed to the effect of mask mandate. According to the final decision tree, we can help government to come up with more reasonable solutions to control the COVID-19.

6. Discussion

6.1 Difficulties encountered:

1. It is difficult to choose the topic, and it is not clear which factors in specific problems in real life are worth using machine learning method to analyze.

2. How to conduct effective data analysis on a single feature and how to eliminate interference factors.

3. Due to the particularity of K-means clustering classification, the number of data points in separate group is uncertain (sometimes the result is this cluster has only one data point), which may sometimes lead to the error of singular matrix in linear regression fitting function.

4. It is difficult to judge whether the linear fitting results are extrapolated or not.

5. When our group used the ID3 algorithm to draw the decision tree, we got no answer. Because after a uniform classification standard, the data cannot be classified by these attributes. However, it doesn't mean that the data has something wrong, it just because the classification standard isn't suitable.

6.2 Solution:

1. We learned from the idea of analyzing Corona-virus cases on the Internet, collected data on the Internet and processed the data with our method.

2. The experimental group and the control group were set, and other factors were excluded.

3. Using the method of exception handling, if there is a singular matrix, the fitting result is zero matrix.

4. Only the slope (infection rate) that is logically associated with the independent variable and the dependent variable is used for calculation and comparison. However, the correlation between the independent variable and the dependent variable remains unproven.

5. Because our group is trying to get relative difference in each attribute of states, we just choose a plausible classification standard to each attribute, and the classification standard haven't strong science basics. As a result, we need to give little change to each attribute for many times until we get the appropriate decision tree. To make the change efficiently, we

use a code to test, if this classification standard is good enough to output a decision tree, which can save time for our research.

6.3 Possible further improvements:

1. Data can be processed by normal distribution model to conduct variance analysis, test for homogeneity function of variance and student's T test.

2. Regression analysis that is closer to biological curves model can be used for data curve fitting, such as J-shaped curves and S-shaped curves, which may have more significant correlation in parametric(feature) analysis.

3. Increase the dimension of data to make the results of K-means clustering grouping more convincing.

4. Do more pre-survey, so that our group can get more reasonable attributes and more scientific classification standard to the decision tree.

5. cooperate with epidemiologists to find solutions to solve the problems which are found by the research.

7. Conclusion

1. Even though some data shows that the new cases per day somehow still increased after mandating mask, we think it indeed had a positive effect on preventing the pandemic. The reason why new cases are still increasing in someplace is due to other unknown factors and we managed to eliminate its effect.

2. According to the efficiency parameter, wearing mask do help to control the COVID-19 in general, so, government should keep this mandate until epidemic is over.

3. According to the decision tree, mask mandates have great effect when the state has high infection rate, which remind the government in high infection rate region to take more strict measures to require people to wear masks.

4. Temperature not only affect the remaining time of the virus, but also can affect the degree of compliance with mandates. If government or companies develop a kind of mask which can alleviate the uncomfortable sense during the hot weather, the effect of mandates can increase.

5. The population density plays different roles in different precondition. We need do more study about that to make government's policies effective.

8. Reference

Data from:

http://seattle-data-dev.s3-website-us-west-2.amazonaws.com/

https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States_by_population_density

https://www.infoplease.com/us/states/state-population-by-rank

https://www.ahd.com/state_statistics.html

https://www.currentresults.com/Weather/US/average-annual-state-temperatures.php