**Factors of COVID-19 Death Rate**

Nitiphong Nirachornkul

Department of Toulouse Graduate School

ADTA 5940 Advanced Data Analytics Capstone Experience

At University of North Texas

Denise Philpot, Ph.D

October 18, 2021

**INTRODUCTION**

This project studies the factors that affect the COVID-19 death rate such as state, age, gender, time period, health condition and the percent of vaccinated to analyze the major factors of the death rate by using data analytical techniques. The project is based on the CRoss Industry Standard Process for Data Mining (CRISP-DM) framework which has five major steps. This dataset is collected from 1 Jan 2020 to 30 Aug 2021.

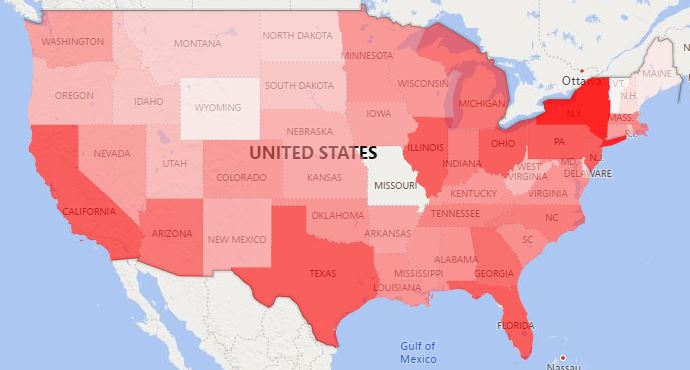
1. **Data understanding**

The data of COVID-19 deaths rate has one numerical variable as the death rate as a predicted variable and three categorical variables as independent variables such as states, health conditions, and age group.

1. **Data preparation**

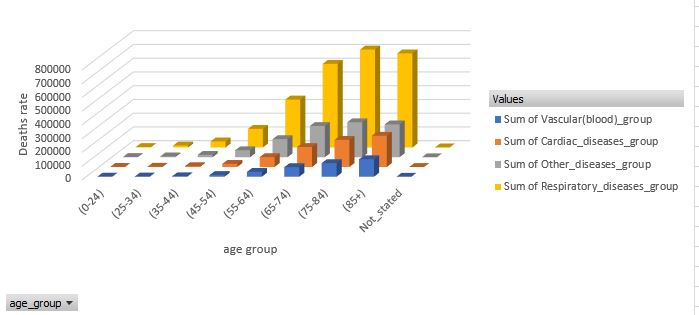
Data preparation steps use Exploratory Data Analysis (EDA) by plotting the chart to analyze the distribution to explore the dataset in different dimensions. However, some datasets are redundant data on several columns such as month, age and year. Some rows of death rate also have missing value because the dataset has been suppressed by NCHS confidentiality standards which need to be cleaned.

**Figure 1 : Map chart**

****

Note: map chart provides the density of covid-19 deaths rate in each state. The red color shows that the death rate is above 500,000. Pink color shows that the death rate is around 70,000 white is less than 70,000. This dataset starts from 1st January 2020 to 30 August 2021.

**Figure 2: Death Rate by Age Group and Health Conditions**

****

Note: The groups of heath conditions are referred from St. Luke Hospital Conditions grouped by organ and body system which can divide into 7 health condition groups such as Blood circulation group, Infectious diseases group, respiratory diseases group, Heart and vascular group, Brain & Nervous System, Endocrine System and Other health conditions group.

1. **Modeling**

Modeling steps start by classifying the outliner factors by classifying models such as KNN and random forest before comparing two models to minimize the number of factors. The process after removing the outlier is splitting the dataset to train and test the model before running on predictive models such as regression models such as linear and stepwise regression before comparing between predictive models.

1. **Evaluation**

The evaluation step by crossing variation is the final analyzing step of this project. Final step of evaluation is summarizing the significant factors to the COVID-19 death rate.

1. **Deployment**

The deployment steps focus on presentations that are illustrated by visualization techniques to provide the most understandable to audiences such as heatmap and barchart to project the predicting trends on the regression models. In addition, this part deploys the ratio of the significant factors such as density of COVID-19 death rate in each area to continuing study in specific locations.

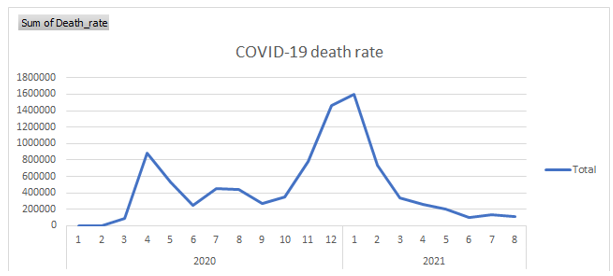
**Background and Significance of Study/Research**

COVID-19 is a disease that is caused by SARS-CoV-2. This virus can spread to the air and infect the human body. The suggestion to slow down infection from the World Health Organization (WHO) is to wear the appropriate filter mask, keep social distances at least 1 meter from each other, frequently wash hands with alcohol, and take vaccine shots following local health organizations direction. The illness of SARS-CoV-2 directly affects the lung. Patients can have mild to moderate illnesses that depend on the patient's health conditions. Patients are at risk of having serious illness or dying from health conditions, some of which are cardiovascular, chronic respiratory disease, disease, diabetes, and cancer (World Health Organization, 2021). Thus, this project will study and prove that health conditions factor is related to the death rate or not. The project studies the mortality rate in the U.S in each state that has a different social environment, different temperature, or weather. The timeline of coronavirus was starting when Dr. Li Wenliang, who is the first coronavirus founder, tried to alarm on December 30th, 2019, about an unknown virus. The first case of coronavirus occurred between 12 December and 29 December 2019. The year 2019 when the first case of COVID pandemic occurred, is the reason for calling the diseases from SARS-CoV-2 as COVID-19. 19 behind the COVID is the year that found the first case. On 20 January, China declared an emergency (Hegarty, 2020). The world’s first COVID-19 death was reported on 9th January 2020 in Wuhan. The first coronavirus patient in the US had been confirmed on 21st January 2020. Lockdown city model to solve the superspreading problem happened for the first time in Wuhan on 23rd January 2020, China. In February 2020, Italy started to use a lockdown model (CNN Editorial Research, 2021).

Currently, COVID-19 epidemic is spreading across the world, causing a higher death rate. COVID-19 mortality rate hurts several nations’ economies. The economic costs can be cash and non-cash such as opportunity cost. The total cost in the Euro zone included non-cash and cash cost is 2.89 billion euro. The top spending cash countries in the Euro zone are Spain that lost 1.07 billion euro, Italy that lost 0.35 billion euro, and Netherlands that lost 0.19 billion euro. COVID-19 death situation hurt Spain’s economy by around 0.11% of GDP (Hanly, Aherm, and Sharp, 2021). The loss of COVID-19 mortality rate is the main reason for this project to study the death rate factors of coronavirus for managing to reduce the cost and death rate.

COVID-19 death rate has a pattern which has peak and low points. In the winter of 2020, COVID death rate peaked and slowed down in the summer similar to the flu season in 1982 to 1983 (NCIRD, 2021). SARS-CoV-2 spreading may rely on the weather and temperature. The date of weather is the one factor that many researchers believe that relates to the spreading rate as the second wave. However, the paper illustrates that COVID infection rate is combined by several factors such as social distancing policy and socio-economic conditions (Engelbrecht & Scholes ,2021).

**Figure 3 : COVID-19 Death Rate**

****

Note: This table shows the COVID-19 mortality rate of the total U.S. from Jan 2020 to August 2021. The dataset in the graph shows the time frame that which period has a high death rate or which period has a low rate.

Mutation of COVID-19 is the new issue. However, mutated coronavirus on the second wave and the third wave, which is in many countries such as India, and other countries in the Eurozone, does not affect the death rate as people believe. According to Effect of mutation and vaccination on spread, severity, and mortality of COVID19 disease article, the mutation COVID-19 cannot be claimed to be significant factors for mortality rate. The main reason, which the article rejects the hypothesis that mutation is the one independent factor of COVID mortality rate, is the death rate of the second and third wave after the virus mutated. The authors of this article use dataset from thirteen countries to prove the correlation between COVID mortality rate and mutation on the second and third wave (Zawbaa, Osama, El‐Gendy and team, 2021).

There are only two major indicators to monitor the COVID situation such as infection rate and mortality rate. Infection rate is not appropriate for this research because not every person is tested. Daily population infection rate needs daily testing on every person. However, not every person can do COVID test everyday with the same testing quality. There are several COVID test types such as Polymerase Chain Reaction (PCR) and Rapid Test Kits (RTKs) (Omori, Mizumoto, & Chowell, 2020). The infection results can be inaccurate because some people are infected but never test. Other infected patients have negative results. Thus, this project focuses on COVID mortality rate under assumption that every COVID death should be recorded and updated.

The consequence of increasing COVID mortality cases affect several dimensions such as economics, mental illness, and conflict. Not only economics is affected by the coronavirus, but also the mental state of people is affected. According to Wang's article, 20 to 25% of the adult population around the world are affected by the pandemic’s mortality rate (Wang, 2020). Mental disorder cases have two major problems such as fear of infection and fear of being killed by the virus. Moreover, depression from COVID leads people to suicide. The pandemic during 2019 to 2020 increases people's stress and leads to anxiety. The COVID-19 epidemic has negatively affected the mental disorder. Suicidality rate increases in the high COVID area because people who have friends or family who have died by the COVID-19 have more stress and higher rate to suicide than other people (Sher, 2020).

Furthermore, a pandemic creates conflict between people. According to the Wang, Gee, and Bahiru article, COVID situations lead some societies to have conflict and lead people to be xenophobic. Merriam-Webster gives the definition of xenophobia to be “fear and hatred of foreigners,” (Merriam-webster, 1828).

The COVID mortality rate relies on many factors such as age range, health condition and location. Some locations are cold. Other locations are hot. Thus, the location should be grouped by states. The three highest death rate states are California at 12.77%, Texas at 10.32% and Florida at 7.12%. These three states which are 30.21% of the death population are not cold temperatures. In conclusion, there are other factors that are more significant than temperature when compared to the death rate that is grouped by states.

According to the dataset, the death rate by age group range shows that the density of age range is dense in the group that is above fifty-five years old. Fifty-five to sixty-four years old group has a mortality rate as 13.05% of COVID death population. Mortality rate of sixty-five years old to seventy-four years old is 23.37%. The mortality rate of seventy-five to eighty-four is 28.09%. Finally, the death rate for those who are above eighty-five is 28.04%. Therefore, the accumulated percentage death rate of people (above fifty-five years old) is 92.91% which is a big amount when compared to younger age.

In addition, the health conditions are also other important factors to the mortality rate that can be classified by several methods. This project focuses the death rate of COVID patients who has health conditions as such Sepsis, Other diseases of the circulatory system, COVID-19 (no health conditions), Influenza and pneumonia, Adult respiratory distress syndrome, Chronic lower respiratory diseases, Respiratory arrest, Respiratory failure, other diseases of the respiratory system, Alzheimer disease, Diabetes mellites, Obesity, Renal failure, Intentional and unintentional injury, Malignant neoplasms, or all other conditions and causes. Twenty-five health conditions can be grouped by human organ into seven groups such as blood circulation, infection diseases, respiratory diseases, heart and vascular, brain & nervous system, endocrine system, and other health conditions (St. Luke's Hospital, 2021).

Sepsis and other diseases of the circulatory system are in the blood circulation group. Death of COVID-19 without conditions and influenza and pneumonia are easy to infect, so influenza and pneumonia and non-conditions are grouped into infectious diseases. Adult respiratory distress syndrome, chronic lower respiratory diseases, chronic lower respiratory diseases, respiratory arrest, respiratory failure, and other diseases of the respiratory system are grouped into the respiratory diseases. There are several diseases in the heart and vascular group, some of which are cardiac arrest, cardiac arrhythmia, heart failure, hypertensive diseases, ischemic heart disease, vascular and unspecified dementia, and cerebrovascular diseases. The fifth group by organ is the brain and nervous system such as Alzheimer disease. The sixth group is the endocrine system that includes diabetes mellitus and obesity. The final group is the group that cannot be grouped by the main organs, which are other health conditions. The other health condition group includes renal failure, intentional and unintentional injury, malignant neoplasms and all other conditions. The fifth group as infection group has the most mortality rate as 41.58%. The infection group includes the death of pure COVID without health conditions. When doing the group summation between the infection diseases, respiratory diseases, heart and vascular diseases, and other health conditions, these groups cover around 90% of the death population.

Heart and vascular disease are the one main health condition that kills COVID patients. Most heart and vascular happens to elders, “a man is as old as his arteries,” Thomas Sydenham said (1624–89). Arteries is the main red blood vassal from the heart. According to Thomas Sydenham, his quote means that when people get old, their blood vassal and other health are old and not working as full function as young people. Heart and vassal health condition starts from kidney is too old to able to filter completely uremia which is the toxic to heart and vascular system. Uremia makes the heart and vascular cells to be old and lead the patients to hyperparathyroidism. In conclusion, the uremia from not having a good functional kidney from elders can lead the elders to have heart and blood vascular conditions (Wanner, Amann, & Shoji, 2016). Thus, elder heart condition might be one of the COVID death rate factors because the dataset shows the highest death rate is from patients who are above fifty-five years old.

Respiratory system is the body system involved in berthing. Thus, respiratory health conditions can be one of the COVID death rate factors. Some patients have lung injury before dying. COVID patients with respiratory health conditions are a very challenging ICU and other medical staff because when patients get inaccurate treatment with the wrong position, patients have more risk. For example, a prone position with a tube is at risk to accidentally remove the tube, pressure ulcer, or other problem (Beshesht, 2021).

Infection diseases health conditions are the disease that can transmit to humans such as COVID-19. Coronavirus that can transfer to the human body is the same as the flu virus and other viruses. The International Committee on Taxonomy of Viruses (ICTV) gives the name of this virus as SAR-CoV-2 and disease name as COVID-19. The Infection diseases not only includes COVID-19 and flu but also includes H1N1 in 2009, H5N1 (influenza A), MERS-CoV, severe acute respiratory syndrome (SARS) in 2002 and other infection diseases (Shereen, Khan, Kazmi, Bashir, & Siddique, 2020).

**Data understanding**

Research scopes on the COVID-19 mortality in 50 states of the U.S. This project focuses on the twenty months period that started from 1st January 2020 to 1st August 2021. Moreover, this project includes every age grouped as 0-24 years old, 25-34 years old, 35-44 years old, 45-54 years old, 55-64 years old, 65-74 years old, 75-84 years old and above 85 years old. This research also includes the group of health conditions of death patients. The data file includes summation of the dataset group.

The first theory that is included in this project is CRoss Industry Standard Process for Data Mining (CRIP-DM) framework. CRIP-DM framework was found in 1996 by five companies such as Integral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation and OHRA. CRIP-DM framework has five main steps to elaborate and solve the problem of organization in simple format for executive team who do not have data background to understand the problems and models. The first main step is understanding business or problems. For example, the problem of this project concerns factors of COVID mortality rate or cause of COVID death to prevent or slow down the mortality rate. Secondly, dataset understanding, and data preparation are the processes to manage missing values, to clean noise of data, to eliminate redundant dataset, and to understand the detail of dataset such as correlation and distribution of dataset. Thirdly, data modeling is the process to analyze the dataset by using statistics models. Fourth step of CRISP-DM is an evaluation that tests the model equation. For example, this project uses cross validation techniques such as bootstrap and k-fold cross validation techniques to evaluate models. Finally, the evaluation results are deployed to audiences in simple format (Fajriansyah, 2021).

**Cleaning data**

The non-COVID death count dataset has some missing values, so this project is designed to use the average of each non-COVID death count condition to fill in missing values because this project tries to minimize the error of the model. Missing values that are changed to be average are less than 10% of the dataset which do not affect the overall dataset.

Moreover, some of COVID death count conditions have missing value, so this project provides the average of each health condition and each state to fill the missing values such as Obesity has 7 missing values, all other cause conditions have 2 missing values, Intentional and unintentional injury poisoning has 3 missing values, neoplasms have 1 missing value, Renal failure has 1 missing value, and sepsis has 1 missing value out of 1021 rows. For example, Obesity has 7 missing values from Wisconsin, Iowa, Pennsylvania, Illinois, Georgia, and Arkansas. The missing value of Obesity from Wisconsin is filled by the obesity condition COVID death case in the Wisconsin area.

**Modeling**

The article applies variable selection techniques such as Mallow’s CP, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) to select the variables for running the predicting model to predict the results of COVID-19 death rate. The prediction models include K-Nearest Neighbors (KNN) and regression models. KNN model is the model to classify the dataset into groups by using the shortest Euclidean distance (Khamar, 2013). The parameters for comparing the results of this project are R-square, and Root Mean Square Error (RMSE). The results of errors have positive and negative values. Another result value that is important to tell the model is R-square which tells correlation of the model and data points. If the R-square is close to one, the model is very appropriate to the dataset (Carmeron & Windmeijer, 1997). The last parameter is Root Average Square Error (RASE) or Root Mean Square Error RMSE. Root Mean Square Error is the parameter that informs the error between predicted results of the model and the actual values (Neill & Hashemi, 2018).

After running the model, the next step is using stepwise techniques to reduce features of the model to avoid overfitting because when the models are too complicated, the models can have a lot of errors when predicting a new dataset. Stepwise technique with forward direction is adding the significant feature one by one and compare Akaike Information Criterion (AIC) to each other which lower AIC model is better than higher AIC model (Akaike, 1978), and Bayesian Information Criterion (BIC) (Akaike, 1978). The AIC and BIC model is a volatility tracking model which tracks standard deviation of dataset. However, BIC weights more data points than AIC.

**Figure 4 Formular**

AIC =

BIC =

n = amount of data point

RSS = Residual Sum Squared

d = number of features or independent variables

= sigma of prediction results

**Evaluating**

Evaluation is testing processes that include cross validation methods such as holdout method, k-fold cross validation, leave-one out cross-validation (LOOCV), and bootstrap method. Cross validation reduces bias by randomly selecting a data subset to train and test the model for finding the range of confidence interval results. K-fold cross validation is the model to split the dataset into K groups which has two types of groups as testing and training dataset. K-fold switches the testing dataset and training dataset until every group is testing the dataset and gets the error results. The model finds average error results for reducing bias of the model (Refaeilzadeh, Tang, & Liu,2009). This project applies k-fold cross validation in the evaluation step.

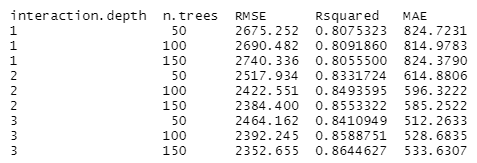
In conclusion, theories involved in this project are CRIP-DM framework, regression, KNN, and are evaluated by k-fold cross validation model. Finally, the final step of the project is exploring the results of each model and finding the best model by visualization technique. The project has four main parts to analyze the dataset: understanding dataset, preparing dataset, modeling, and evaluating. The final step is about exploring the dataset and the results to the audiences.

**The results of the modeling**

The first prediction model of this project is KNN. K-Nearest Neighbors (KNN) is the model that separates the similar data factors into the same group by the Euclidean distance. The first step, the model splits the dataset into two groups such as training dataset and validating dataset. KNN model is the model to predict the COVID deaths rate and compare between real deaths rate and predicted values. The results of validating separate the 38 independent variables into 9 groups which is the best result when compared to other separating results. The result of 9 groups in validating processes has the lowest RASE as 3885.62 and highest R-square as 0.5234. This model is designed to compare 2 data groups to 40 data groups.

The second predict model is a boosted tree model which sets the target variable to be COVID mortality rate and other factors are independent factors. This model runs with 70% of the dataset to train the model.

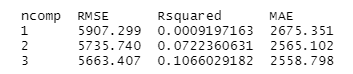
**Figure 5: Boosted Tree Result**

****

Note: Tuning parameter 'shrinkage' was held constant at a value of 0.1. Tuning parameter was held constant at a value of 10 RMSE and was used to select the optimal model using the smallest value. The final values used for the model were n.trees = 150, interaction.depth = 3, and shrinkage = 0.1. RMSE is 2352.66, and the best RSquare is 86.45%.

The fourth model is the partial least squares model. The partial least squares model is the prediction model that compares multiple linear models to find the least error between linear models. The least error has RMSE as 5663.407 and R-squared as 0.1066.

**Figure 6: Partial Least Squares Results**



Note: the partial least squares model creates multiple linear equations and compares them with each model.

The last prediction model is linear regression. The linear regression is the simplest model which finds correlation and fills in linear equations. The result of linear regression has RMSE as 9.7646 and R-squared as 0.999982.

**Table 1: Summary Results**

RMSE R-sqared

KNN 3885.61992801173 0.523426058482371

linear 10.4088058431026 0.999991217659444

Partial least squares 5357.21317716596 0.112276372798351

Boosted tree 2352.65477407941 0.864462661115263

Note: the linear model is the best model when compared to the RMSE which is the lowest and highest R-squared.

However, other models besides linear are underperforming. This project focuses on classifying the factors that affect COVID mortality rate which can use hypostasis testing in the best model which is linear model. The COVID death rate is the summation of health conditions with COVID death rate results, so the first hypostasis testing shows that COVID death rate and health conditions with COVID death rate have a high correlation. Thus, the second hypostasis testing model excluded COVID death with health conditions.

**Table 2: Significant Independent Variables**

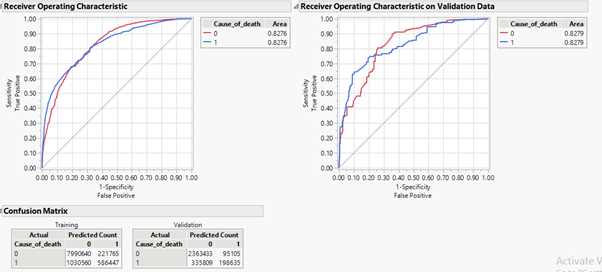
| **Significant Independent Variables** | **P value of (student-t)** |
| --- | --- |
| California | 7.48 x 10-7 |
| Month | 0.050228 |
| Florida | 2.91 x 10-5 |
| Illinois | 1.65 x 10-5 |
| Maryland | 0.008761 |
| Michigan | 0.000125 |
| New Jersey | 0.00982 |
| New York | 6.17 x 10-10 |
| Ohio | 0.000294 |
| Oklahoma | 0.0194 |
| Pennsylvania | 1.58 x 10-5 |
| Wisconsin | 0.029178 |
| Nature Death Rate | < 2 x 10-16 |
| Neoplasms | < 2 x 10-16 |
| Heart Diseases | 3.62 x 10-11 |
| Unclassified Death | 0.000292 |

Note: significant level is 0.05

**Normal Logistics Regression**

Logistics Regression is one of the predictions of the model in binomial form. This project merges the dataset between the COVID death rate and the nature death rate by health conditions such as the heart and vascular group, some of which are cardiac arrest, cardiac arrhythmia, heart failure, hypertensive diseases, ischemic heart disease, vascular and unspecified dementia, and cerebrovascular diseases. The patients who die because of COVID-19 and heath conditions written by 1, and other patients who are non-COVID deaths written by 0.

**Figure 7: Logistics Regression**

****

Note: the results of the logistic regression model shows that the accuracy of the model is around 80% to predict between COVID death and non-COVID death. The results show that the after validation method predicts 293,740 death cases because of COVID-19, but the model correctly predicts 198,635 death cases.

**Stepwise Logistics Regression**

**References**

Akaike A (1969) Statistical predictor identification. Annal of the Institute of Statistical Mathematics 22, pp. 203-217.

Akaike A (1978) Bayesian analysis of the minimum AIC procedure. Annals of the institute of Statistical Mathematics 30, part A, pp. 9-14.

Cameron, C., & Windmeijer, F. A. G. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, *77*(2), 329–342. https://doi.org/10.1016/s0304-4076(96)01818-0

Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases (NCIRD). (2021, September 28). *Flu season*. Centers for Disease Control and Prevention. Retrieved October 11, 2021, from<https://www.cdc.gov/flu/about/season/flu-season.htm>.

CNN Editorial Research. (2021, October 7). *COVID-19 pandemic timeline fast facts*. CNN. Retrieved October 11, 2021, from<https://www.cnn.com/2021/08/09/health/COVID-19-pandemic-timeline-fast-facts/index.html>.

Engelbrecht, F. A., & Scholes, R. J. (2021). Test for covid-19 seasonality and the risk of second waves. *One Health*, *12*, 100202–100202.<https://doi.org/10.1016/j.onehlt.2020.100202>

Fajriansyah, G. (2021). Analisis Daftar pemilih Tetap Pada Hasil Rekapitulasi KPU berdasarkan usia menggunakan algoritma K-means (Studi Kasus : Kota Bandar Lampung). *Analisis Daftar Pemilih Tetap Pada Hasil Rekapitulasi KPU Berdasarkan Usia Menggunakan Algoritma K-Means (Studi Kasus : Kota Bandar Lampung)* , *15*(1), 39–53.<https://doi.org/10.23960/elc.v15n1.2147>

Hegarty, S. (2020, February 6). *The Chinese doctor who tried to warn others about coronavirus*. BBC News. Retrieved October 11, 2021, from<https://www.bbc.com/news/world-asia-china-51364382>.

Hanly, P., Ahern, M., Sharp, L. *et al.* The cost of lost productivity due to premature mortality associated with COVID-19: a Pan-European study. *Eur J Health Econ* (2021).<https://doi.org/10.1007/s10198-021-01351-8>

Ke, X., & Hsiao, C. (2021). Economic impact of the most drastic lockdown during COVID‐19 pandemic—the experience of Hubei, China. *Journal of Applied Econometrics*.<https://doi.org/10.1002/jae.2871>

Khamar, K. (2013). Short text classification using KNN based on distance function. *International Journal of Advanced Research in Computer and Communication Engineering*, *2*(4), 1916-1919.

Merriam-Webster. (1828). *Xenophobia*. Merriam-Webster. Retrieved October 12, 2021, from<https://www.merriam-webster.com/dictionary/xenophobia>.

Neill, S. P., & Hashemi, M. R. (2018). Ocean modelling for Resource Characterization. *Fundamentals of Ocean Renewable Energy*, 193–235. https://doi.org/10.1016/b978-0-12-810448-4.00008-2

Omori, R., Mizumoto, K., & Chowell, G. (2020). Changes in testing rates could mask the novel coronavirus disease (COVID-19) growth rate. *International Journal of Infectious Diseases*, *94*, 116–118.<https://doi.org/10.1016/j.ijid.2020.04.021>

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. *Encyclopedia of database systems*, *5*, 532-538.

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). Covid-19 infection: Emergence, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*, *24*, 91–98. https://doi.org/10.1016/j.jare.2020.03.005

St. Luke's Hospital. (2021). *Complementary and alternative medicine*. Conditions grouped by organ and body system | Complementary and Alternative Medicine. Retrieved October 12, 2021, from<https://www.stlukes-stl.com/health-content/medicine/33/002350.htm>.

Thomas Sydenham Quotes. (n.d.). BrainyQuote.com. Retrieved October 15, 2021, from BrainyQuote.com Web site:<https://www.brainyquote.com/quotes/thomas_sydenham_176953>

Wang, D., Gee, G.C., Bahiru, E. *et al.* Asian-Americans and Pacific Islanders in COVID-19: Emerging Disparities Amid Discrimination. *J GEN INTERN MED* 35, 3685–3688 (2020).<https://doi.org/10.1007/s11606-020-06264-5>

Wang, Q. Q., Xu, R., & Volkow, N. D. (2020). Increased risk of COVID-19 infection and mortality in people with mental disorders: analysis from electronic health records in the United States. *World Psychiatry*, *20*(1), 124–130.<https://doi.org/10.1002/wps.20806>

Wanner, C., Amann, K., & Shoji, T. (2016). The heart and vascular system in Dialysis. *The Lancet*, *388*(10041), 276–284.<https://doi.org/10.1016/s0140-6736(16)30508-6>

World Health Organization. (2021). Coronavirus disease (COVID-19). Retrieved October 12, 2021, from<https://www.who.int/health-topics/coronavirus#tab=tab_1>.

Zawbaa, H. M., Osama, H., El‐Gendy, A., Saeed, H., Harb, H. S., Madney, Y. M., Abdelrahman, M., Mohsen, M., Ali, A. M., Nicola, M., Elgendy, M. O., Ibrahim, I. A., & Abdelrahim, M. E. (2021). Effect of mutation and vaccination on spread, severity, and mortality of COVID‐19 disease. *Journal of Medical Virology*, 6–6.<https://doi.org/10.1002/jmv.27293>