Predicting the Unpredictable: How Machine Learning and Data Visualization are Helping to Tackle the Coronavirus Outbreak in the United States

Temitope, O. Awodiji
Department of Information Security, University of Cumberlands, Kentucky, USA
tawodiji79374@ucumberlands.edu

Aderonke, D. Tosin-Amos
Department of Business Administration, California Miramar University, San Diego, CA.
atosin-amos@student.calm.edu

Femi Ayoola
Jackson State University, Jackson, MS.
femi.ayoola@jsums.edu

John Owoyemi
University of the Cumberlands, Williamsburg, Kentucky
owoyemi08059@ucumberlands.edu

***Abstract*—In the context of global health, it is crucial to both stop COVID-19 from spreading and completely eradicate it. The United States of America is one of the nations in the globe that has suffered from the pandemic's outbreak. To inform readers and stop the pandemic from spreading, this study intends to establish a framework for effectively using machine learning and data visualization. To understand how predictions can be generated from visualized data and how to prepare for the pandemic's effects, this study investigates the potential and opportunities in machine learning. The combination of domain knowledge per introspective theoretical insights, current data sources, and the efficient application of statistical tools results in a wealth of advanced data quality and application options.**

This research utilizes three multi-variate machine learning methods, such as Support vector machines (SVM), neural networks (NN), and Bayesian networks, to anticipate and stop the spread of the epidemic. (BN). The numbers of infections, recoveries, deaths, and vaccinations within three months are tested using these algorithms. The performance of each machine learning method was evaluated using three statistical parameters: Root Means Squared Error (RMSE), Mean Absolute Error (MAE), and Explained Variance Score (EVS). The findings showed that Bayesian Network (BN) and Neural Network (NN) performed better as predictive mechanisms than other models, and the overall finding suggested that while the spread of the viral disease might not slow down, the vaccination rate might be slowed down, and the number of fatalities will keep reducing.

Key words–machine learning; support vector machine; Bayesian network; neural network; data source; vaccination; disease.

I. INTRODUCTION

As of April 2nd, 2023, there were 102,697,566 confirmed instances of the viral sickness in the United States, with 1,117,054 deaths, the highest number ever for any nation [3]. The pandemic is listed as the tragedy with the highest death toll in the United States as of 2021 [4]; in 2020, it was the leading cause of death in the country, trailing only heart disease and cancer [5]. For Hispanic and Latino Americans, African Americans, and White Americans, the U.S. life expectancy decreased from 2019 to 2020 by 3 years, 2.9 years, and 1.2 years, respectively [6]. These impacts persisted as life expectancy continued to decline from 2020 to 2021 [7] and COVID-19-related deaths in the United States in 2021 outnumbered those in 2020 [8].

The first case of this virus in the United States was reported on January 13, 2020, and on January 31, 2020, the US government declared the epidemic a public health emergency. Over 1.1 million deaths have been recorded in the United States of America since the first known deaths occurred in February. Even though the government earmarked $8.3 billion to combat the epidemic in 2020, the infection rate appears to be high [9]. The death rate may give the impression that no attempts have been taken to stop the spread and effects of COVID-19 in America.

The epidemic caused the stock market to plummet and irreparably damaged the country's economy. After the economic cycle peaked in February 2020, it caused a recession in the US [9]. The economy shrank by 4.8% in 2020 from January through March, while the unemployment rate increased to 14.7% in April. With a record yearly decrease rate of 32.9, the impact of COVID-19 on the US GDP cannot be ignored [10]. The COVID-19 pandemic's negative economic effects and widespread unemployment did give rise to fears of a mass eviction crisis, which have since materialized. According to an analysis by the Aspen Institute, 30–40 million people could face eviction by the end of

2020 [11]. In conclusion, the pandemic's worsening effects continue to affect not just the American economy but the entire world, and we must be ready with adequate foresight tools and protective measures to help with recovery and stop the infection from spreading further. The Bureau of Statistics [12] depiction of the pandemic's effects on several economic indicators in 2020 is shown below.

Variables	Feb	Mar	April	May	June	July	Aug	Sept	Oct	Nov
Jobs monthly change	251	−1,373	−20,787	2,699	4,800	1,780	1,371	661	653	256
Jobs, level (000s)	152,463	151,090	130,303	133,002	137,802	139,582	140,914	141,720	142,373	142,629
Unemployment rate %	3.5%	4.4%	14.7%	13.3%	11.1%	10.2%	8.4%	7.9%	6.9%	6.7%
Number of unemployed (millions)	5.8	7.1	23.1	21.0	17.8	16.3	13.6	12.6	11.1	10.7
Employment to population ratio %, age 25–54	80.5%	79.6%	69.7%	71.4%	73.5%	73.8%	75.3%	75.0%	76.0%	76.0%
Inflation rate % (CPI-All)	2.3%	1.5%	0.4%	0.2%	0.7%	1.0%	———	———	———	——–
Stock market S&P 500 (avg. level)	3,277	2,652	2,762	2,920	3,105	3,230	3,392	3,380	3,270	3,694
Debt - held by public ($trillion)	17.4	17.7	19.1	19.9	20.5	20.6	20.8	21.0	21.2	21.3

Table 1. Impact of the Pandemic on various economic variables in 2020

According to the aforementioned, the advent of COVID-19 created a new shift in the health sector and it had a significant impact outside of the sphere of health. More than the incidence of the disease itself were the effects felt by patients and people in general. It permanently altered the prior social, political, and cultural structures of humanity. COVID-19 is a continuing threat to the world, and as such, it requires sufficient attention.

II. Machine Learning

In recent years, Machine Learning (ML) has established itself as a technology that can resolve extremely challenging real-world situations [13]. The majority of activities, including natural language processing, autonomous vehicles, robots, weather forecasting, and medicine, use machine learning and artificial intelligence [14]. The methods used by machine learning algorithms and conventional epidemiological models to forecast COVID-19 spread and control are different. ML employs the process of learning from mistakes and fixing them as a result of those mistakes. One of the most crucial applications of ML is prediction.

However, the growing body of information on many facets of the pandemic can be confusing, and the information about COVID-19 has occasionally been misinterpreted in the media [15], which may make it more challenging to stop the epidemic. Therefore, one of the goals of this essay is to provide an accessible illustration that will aid in the battle against the pandemic's spread within the United States. ML is important for pandemic prediction since, in contrast to the conventional epidemiological method, it will leverage predictive factors such as illness severity, recovery rate, dissemination, and fatality rate. The findings presented here appear to provide a thorough understanding of the illness and support efforts to combat it. This document, which acts as a scientific and social conduit, is set up so that readers may gain an understanding of the pandemic by using ML to deliver accurate and up-to-date data as well as significant news stories for COVID-19 in the US. This study will utilize a reliable cluster of data from OurWorldInDate.org. The major goal is to make it easier and more intuitive for the general population to grasp the information linked to the disease so they can decide how to respond to the pandemic, which is still an ongoing scourge.

III. Related Works

1. Machine Learning and the Traditional Epidemiological Models

A kind of artificial intelligence known as "machine learning" and "big data analytics" enables computer systems to utilize previous data to anticipate future outcomes of events or studies with high accuracy without having been expressly programmed to do so. Big data analytics is the process of gathering, storing, and analyzing enormous volumes of data to draw conclusions and make wise judgments [16]. Machine learning is a powerful tool for modeling real-world issues since it can be utilized in data analysis to uncover patterns and generate predictions [17]. According to research by Arogyaswamy [18], several of the top technological firms, like Facebook and Google, are using machine learning more in their daily operations.

The Susceptible-Infected-Recovered (SIR) model is one example of a Traditional Epidemiological

Model (TEM), which is based on mathematical equations that include data on the characteristics of a chosen event (in this case, COVID-19) and the population's behavior. The alternative method uses ML algorithms to discover links and patterns without having to explicitly train them to do so. The purpose of this study is to decide the future of COVID-19 in the United States by comparing the two ways based on their qualities, even if each strategy has its advantages, disadvantages, and strengths. The following are some of the models' shared characteristics:

1. Flexibility and speed: Traditional epidemiological models need a lot of time and effort to alter, but machine learning algorithms can swiftly adapt to changing conditions and integrate new data [19]. As a result, machine learning algorithms are more adaptable and effective in circumstances that change quickly, like the COVID epidemic.
2. Complexity: Machine learning algorithms are capable of processing vast, intricate datasets that are challenging for conventional epidemiological models to manage. As a result, traditional models may miss certain patterns and linkages that machine learning algorithms may use to better accurately anticipate how the pandemic will develop in the future.
3. Interpretability: Compared to machine learning algorithms, which can be opaque and challenging to grasp how the algorithm generated its predictions, traditional epidemiological models are frequently more clear and simpler to interpret. This is crucial so that policymakers and public health professionals may base their decisions on the findings in an educated manner.
4. Accuracy: Depending on the type and quantity of data available as well as the specific models and methodologies utilized, machine learning algorithms and conventional epidemiological models may be useful in forecasting the spread of the virus. Traditional epidemiological models may be more accurate in some situations, while machine learning methods may perform better in others. However, overall, ML results are typically more accurate than TEM, except for interpretability, which appears to be technical for ML method output.

Predictions made using both models together would be nearly flawless. Date, count, recuperation, and death are only a few of the input parameters that will be used to construct this forecast. The following list classifies machine learning algorithms:

1. Supervised algorithms: Supervised ML algorithms apply previously learned lessons to new data. Supervised learning systems use labeled data to create predictions. The algorithms start by learning from a collection of datasets that were previously known. Even further, it uses the established set of principles to predict the data for unknowable values. After sufficient training, the system is capable of producing outputs for new input. The learning algorithms compare the predictions to the actual output. To accommodate for the output's errors, further changes are applied [20].
2. Semi-supervised algorithms: Between supervised and unsupervised learning, semi-supervised machine learning approaches sit in the center. They use both labeled and unlabeled data for training. Over time, these techniques help systems operate better. Additional unlabeled data is commonly provided to semi-supervised algorithms. Semi-supervised learning is used when a competent intervention is required. Unlabeled data are not required, hence no more expert activities are required [21].
3. Unsupervised algorithms: These machine-learning techniques do not require labeled data. The unsupervised algorithm receives just unlabeled data as input. There is no right or incorrect way to classify anything. The computer software analyzes the data and groups it [22].

IV. Methods

1. Overview

One must learn about the virus to forecast its patterns, and doing so requires the generation of enormous amounts of data. State and national statistics, growth rate variation, COVID-19 forecasts, lab data, disparities and at-risk groups, economic data, mobility, lab data, and mortality are the major categories into which this data is divided. This study will utilize machine learning algorithms to develop specific approaches to make predictions in test and real environments by using a portion of the data given above [23]. There are still some underlying obstacles that this approach may encounter, but it is clear that its advantages surpass its challenges. As we have seen since 2019 and beyond, the pandemic has challenged the US medical prowess and socioeconomic power. Explained below are the ML algorithms employed in this study.

Support vector Machines (SVM): is a type of supervised machine learning algorithm that is commonly used for classification and regression analysis [26]. The SVM algorithm can be used for both linear and nonlinear classification and regression tasks by using different types of kernels, which allow the algorithm to map the data into a higher-dimensional space where a linear boundary can be found [27]. SVMs are robust to noise, handle high-dimensional data well, and can be used for both binary and multi-class classification tasks which are the basis of predicting the trends of COVID-19 [28].

Neural Network (NN): an NN is a kind of machine learning algorithm that is based on how the human brain works and is structured. It is made up of 'neurons,' or linked nodes, which cooperate to process and interpret data. Each layer in a neural network serves a particular purpose in the learning process. The output layer generates a prediction or a judgment using the data from the input layer [29]. The data processing is done by the hidden layers in between, who convert the incoming data into a format that the output layer can utilize [30].

Bayesian Networks (BN): in applications including decision-making, diagnosis, and risk assessment, where uncertainty and probabilistic reasoning are crucial, Bayesian Networks (BN) are frequently utilized. The edges between the nodes in a BN indicate the conditional dependencies between the variables, and each node in a BN represents a random variable [31]. A parent node directly impacts its offspring nodes since the direction of the edges denotes the direction of causation. Each node has a conditional probability distribution attached to it that describes the likelihood of the node given its parents.

2. ML Prediction Models

SVM, NN, and BN are the three regressions that have been employed in this work to project COVID-19 trends. We used the ML Perceptron Regressor (MLP) to equal the NN [32]. Limited-Memory Broyden-Fletcher-Goldfab-Shannon (LM-BFGS) algorithms are used in the MLP model to maximize the squared loss. Every partial derivative step of the loss function about the model parameters that are computed to update the parameters is where the MLP repeatedly operates. This work used the iterative idea to acquire the partial derivatives of the loss function about the model's parameters to compute the updated parameter values [33–36]. We employed the following parameters in our model, which was modeled after ref [37]:

a. *Alpha = 0.0001, hidden layer sizes = (30, 30), random state = 35, learning rate = constant, solver = lbfgs, activation = relu in the event of recoveries.*
b. *in the event where the verified alpha is 0.001, the hidden layer sizes are (100, 60), the learning rate is constant, the solver is lbfgs, and the activation is relu.*
c. *In the event of fatalities, the parameters are as follows: alpha = 0.001, hidden layer sizes = (500, 120), random state = 25, learning rate = constant, solver = adaptive, activation = relu.*

The "kernel mechanism" was employed by the SV Model following the goal of this study. The translation of input data into anticipated output outcomes was aided by the kernel mechanism. Converting the study's input (x) to an n-dimensional space aids in the resolution of any faulty linear regression method. (z). Due to the non-linear behavioral method that this study suggested using, this initial solution is necessary. The following SVM parameter was used in the paper:

a) *In the event of recovery, shrinking=True; kernel='poly'; gamma=0.01; epsilon; 1; degree; 4; C;*
b) *In the event where (shrinking=False, kernel='poly, gamma=0.01, epsilon=0.01, degree=3, C=0.1) infection is confirmed.*
c) *In the event where (Shrinking=True, Kernel='Poly', Gamma=0.1, Epsilon=0.01, Degree=2, C=0.01;) then it is fatal.*

Finally, by combining Sparse Bayesian Learning with the Relevance Vector Machine, the BN was used to build the models that needed parameter updates [38]. The Bayesian modeling framework has demonstrated its adaptability and near-perfection prediction of a regression approach in terms of a hierarchical data structure [38].

V. Methodology

A user-friendly web app system built on machine learning frameworks serves as the suggested system. The goal of this system is to use SVM, BN, and NN to recognize, understand, and forecast trends in COVID-19 cases in the US. The interface for the system is shown below.

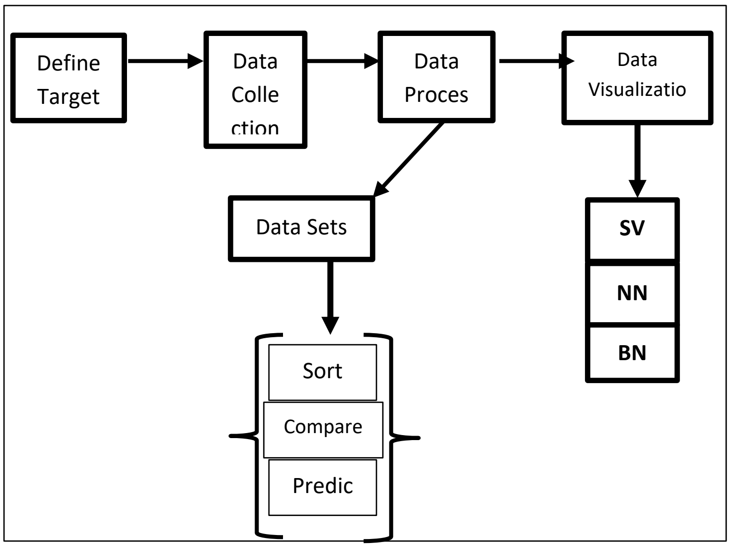

Figure 1. Proposed System Operation Interface

Target Data

The first step in the illustration above is to specify the goal data; in this example, our target data are COVID-19 confirmed cases, populations with and without vaccinations, recoveries, and fatality rate.

Data Collection

The gathering of the target data comes after the goal specification as the next course of action. Data for this suggested system will be acquired from The News York Times' Coronavirus Data in the United States Dataset [40] and Our World in Data's COVID-19 Github Dataset [39]. The former includes regularly updated and aggregated government data. It includes information on all 50 states. The databases for these states provide information on daily cases, recoveries, fatalities, and immunizations. The latter offers US-specific information that offers thorough state and county-level case data. The suggested system will be connected to both data sources so that the datasets may be automatically decoded every day from 12 am to 5 am. The fetched dataset during this time will serve as the basis for processing.

Data Processing

Many machine learning techniques require the standardization of a dataset to achieve better results since the algorithm may perform poorly if the individual characteristics are not standard, regularly distributed data. Before being fed into the machine learning method, the numerical data from the COVID-19 cases dataset is normalized using the StandardScaler from the sklearn Python library.

Datasets

The following structure, which includes the fields ID, day, month, year, state, new, recovered, and died, is appropriate for processing the data file. Pre-processing files is the process of preparing data for operational use. The specification of this format is what may be referred to as the main dataset for the prediction. The dataset will be sent either manually or automatically using a script. Human involvement will be used to prepare the dataset for the present scope. Below is a summary of the fields:

ID: a unique identification, only for the record.
Month: that year's month, as well as that month's day (formatted as Jan/01).
Year: the day, month, and year (formatted as dd/mm/yyyy).
State: name of the state.
Infected: total new patients who have been infected as of that date and in that state.
Recovery: total patients that were recovered as of that date (recovered).
Fatality: total number of deceased patients.

Datasets Comparison and Prediction

Due to the risk to human life, this study aims to discover a novel technique to restrict the transmission of the Coronavirus. Globally, both the incidence of human irresponsibility and infection are rising daily. This study's goal is to forecast confirmed cases, recoveries, and fatalities in the upcoming days to help manage the pandemic crisis. The dataset utilized in the study contains regular time series data on the number of confirmed cases, fatalities, and recoveries during the days since the pandemic began until July 18, 2022. The number of confirmed cases, recoveries, and fatalities will be predicted. The system will use the three machine learning algorithms—SVM, BN, and NN—mentioned above to perform prediction and select the algorithm that produces the best results out of the three using the Root Means Squared Error (RMSE), Mean Absolute Error (MAE), and Explained Variance Score (EVS) functions from the *sklearn* package to calculate the error rate, performance, and scores, respectively. The system will divide the data into training and testing sets using the train-test-split (TTS) function from the *sklearn* package after preprocessing the data. The models are then trained, and predictions are made, using the fit (f) and predict (p) functions. The database holds the outcomes for subsequent use. The dataset (281) was divided into a training set (221 days) and a testing set in the following stage (60 days). These models were trained using trends from confirmed instances of recoveries and deaths. EVS, RMSE, and MAE are only a few of the metrics that make up the evaluation matrix. The suggested strategy utilized in this article is shown in the following manner:

ISSN: 2458-9403
Vol. 10 Issue 4, April - 2023

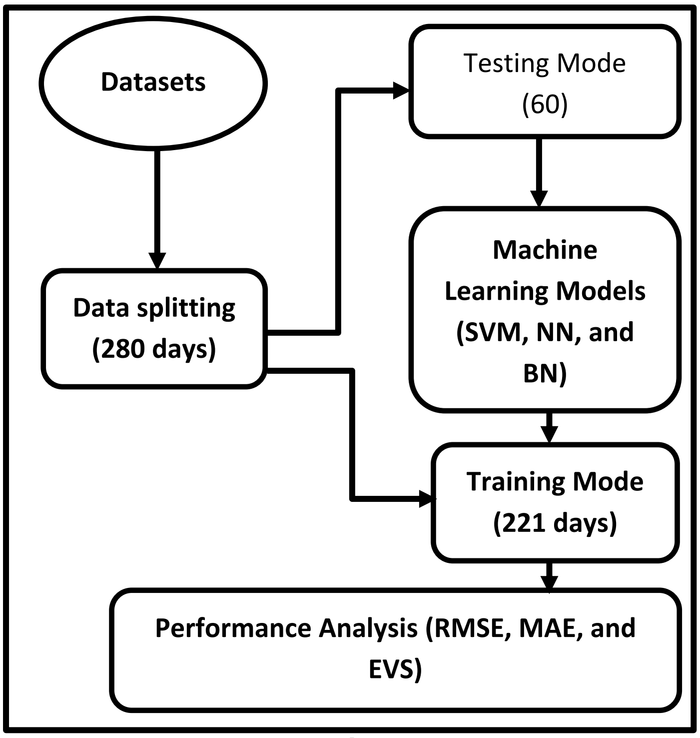

Figure 2. ML System Approach

Data Visualization

As demonstrated in Figures 3, 4, 5, and 6 below where the log values are displayed from Figures 7 to 10 all data before March 18, 2023, were utilized to construct the anticipated findings. The daily instances from the US are shown in Figs. 10 through 12. The numbers represented data on a moving average over a period of 10 days, which can be calculated using the equation:

$$\frac{x_t + x_{t-1} + x_{t-2} + (a) + x_{M-(t-1)}}{M}$$

Where (M) represents the absolute mean window. The leveling in the absolute mean (M) would aid the accuracy of the prediction.

VI. Visualization and Prediction of COVID-19 Using Machine Learning

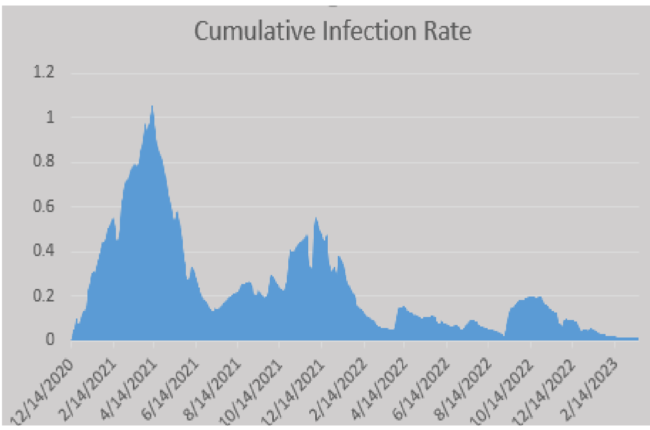

Figure 3. Cumulative Infection Rate

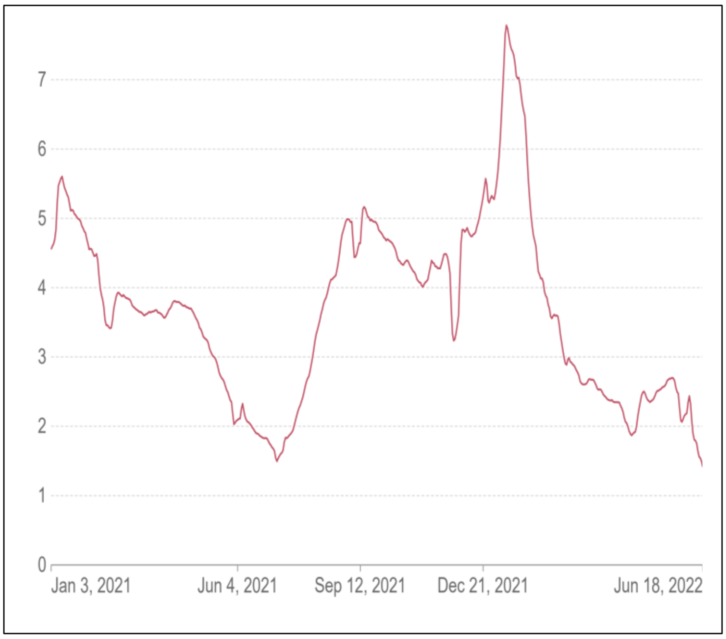

Figure 4. Average Daily Tests per Thousand People

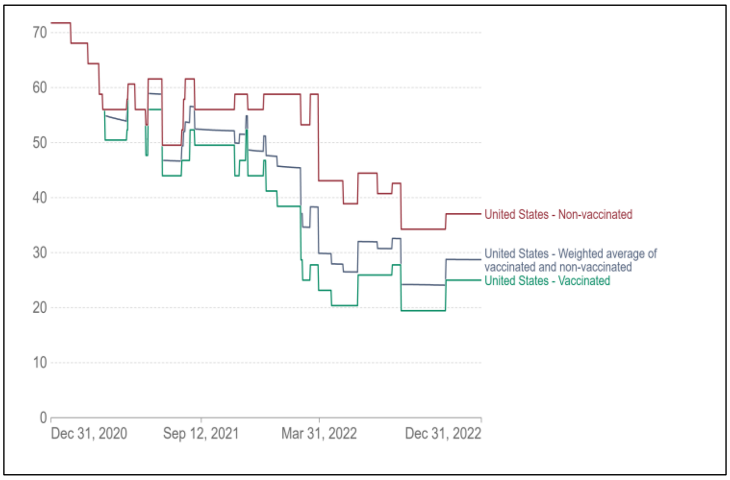

Figure 5. Vaccination Index

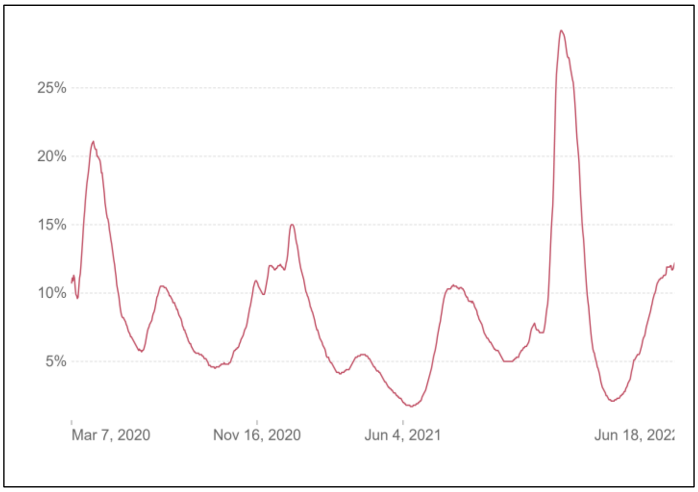

Figure 6. Daily Positive Results

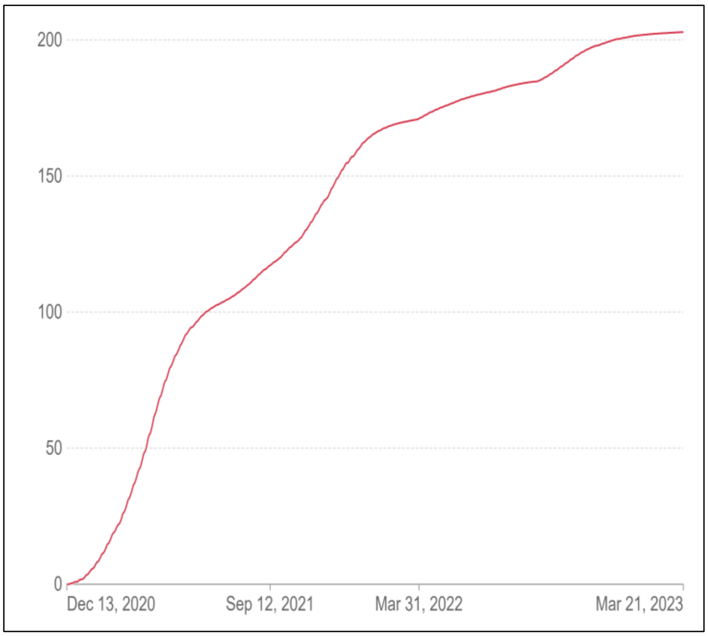

Figure 7. Vaccination Doses per 100 People

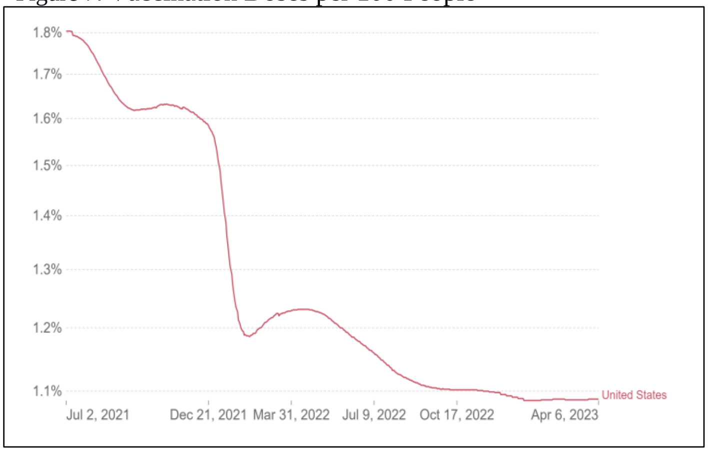

Figure 8. Mortality Rate

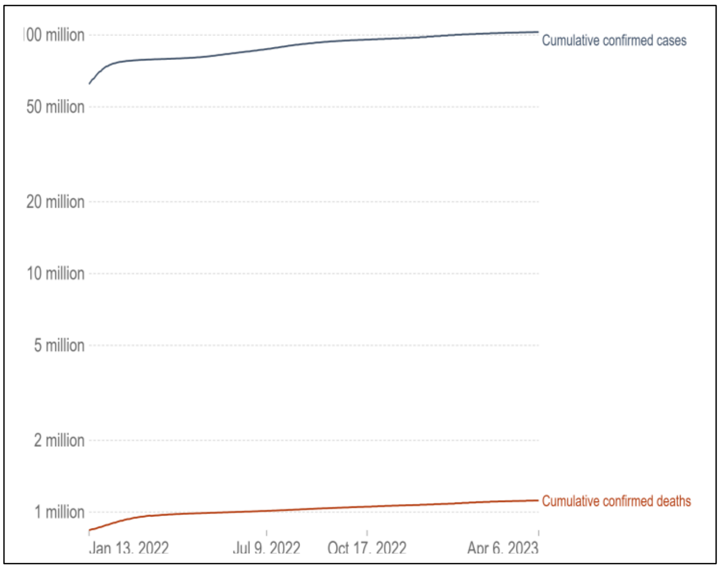

Figure 9. Cumulative Deaths and Cases

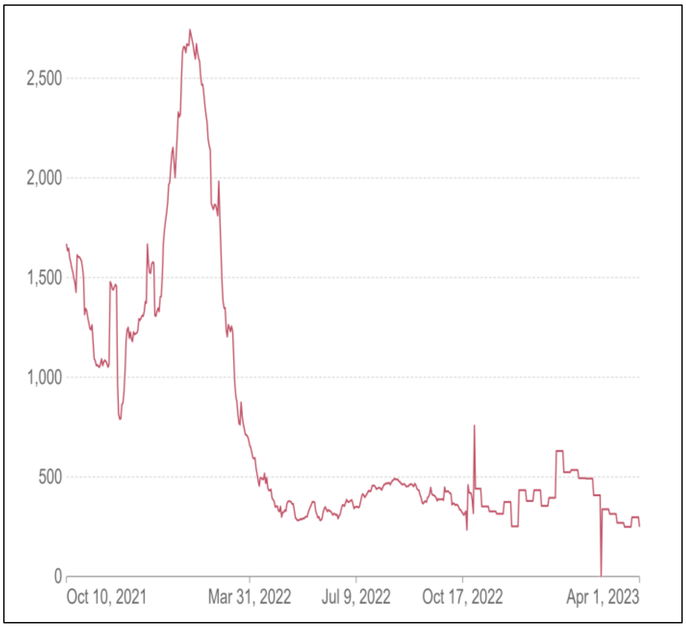

Figure 10. Daily Confirmed Deaths

VII. Predictions and Result

Three machine learning models (SVM, NN, and BN) have been used to forecast the number of confirmed cases, deaths, vaccination, and recoveries. The number of people infected with COVID-19 is not known across the US because not all cases are reported especially from the counties. This paper has attempted to know the number of infected, recovered, vaccinated, reported, confirmed, and death cases in the US by employing the ML algorithms above in the coming days.

When compared to the COVID-19 case trend, the impact of the newly reported instances has significantly dropped and has given the capacity to live comfortably. Utilizing these prediction tools has helped us to be prepared to handle these difficulties (death and vaccination). The surprise factor has given COVID-19 the advantage above us particularly during the second wave, allowing the states to make hasty and occasionally impulsive choices for the benefit of the residents. When we think back to the two waves that affected the United States, we can see that the preparation for the pandemic was lacking. Now that these technologies are accessible, the nation has been more than prepared for future unforeseen pandemics or tsunamis.

It's crucial to test our model with fresh information to prevent overfitting to the training set. We can't do this with the test set since we'll end up choosing the parameters that perform best on the test data, not necessarily the ones that generalize best. However, it might be good to evaluate our model while we're constructing it to discover the optimal parameters. To evaluate the model while it is still being developed and modified, we produce a third collection of data known as the validation set. In a typical train/test/validation split, 60% of the data would be utilized for training, 20% for validation, and 20% for testing. Below is a graphical accuracy representation of selected algorithms:

VIII. Ethical Consideration

Data privacy is a crucial ethical issue when using machine learning algorithms and data visualization approaches. It takes a lot of data to train these algorithms and create useful models. However, this data frequently contains private information like medical records, location information, and personal data. We must make sure that this information is gathered and stored responsibly and securely with the proper entities' agreement. Best practices for data sharing should be followed, as well as moral and legal requirements. It is significant to stress that during emergencies, such as a pandemic, data privacy may temporarily be infringed for public health concerns; however, this should be appropriate, and transparent, and the data should, wherever feasible, be anonymized.

Data bias is a crucial ethical point to address because machine learning algorithms can only be as effective as the data they are trained on, biased data will result in biased predictions and judgments. Numerous factors, including sampling bias, selection bias, and algorithmic bias, can lead to prejudice. Decisions based on race, gender, age, or other criteria may have an uneven influence as a result of this. To maintain a fair and equitable process, it is crucial to spot any biases in data and algorithms, minimize them, and engage a variety of stakeholders in their creation. Usually, security events will be related to some other actions: illegitimate access to data confidentiality damage, injury to the integrity of knowledge, and loss of data accessibility (Discover) [41]. Loss of privacy of data, creating them accessible to others without a right of access is not visible within the database and does not need changes without the deductible database [42].

The possible abuse of data visualization and machine learning capabilities is the last ethical consideration. These tools may be used to follow people, observe their actions, and restrict their movement, which raises concerns about privacy, autonomy, and human rights. It's crucial to make sure that these technologies are utilized ethically, openly, and with the proper supervision and accountability systems in place. A variety of stakeholders, such as public health experts, legislators, civil society groups, and impacted communities, should be included in these procedures.

In sum, the use of data visualization tools and machine learning algorithms to forecast and stop the spread of the virus poses significant ethical questions about data privacy, data bias, and possible technological abuse. Addressing these issues in a fair, responsible, and equitable manner will take teamwork from a variety of stakeholders.

IX. Conclusion

In conclusion, machine learning and data visualization approaches are essential for comprehending and studying the COVID-19 epidemic in America. The transmission of the virus has been examined, hotspots have been located, and future trends have been predicted using Bayesian Network, Neural Network, and Support Vector Machine (SVM) methods. These models have shown to be successful at offering perceptions and suggestions for public health actions.

To uncover risk factors and viable treatments, causal links between variables can be examined using Bayesian networks. Neural networks can recognize intricate patterns in data, which may be used to forecast the number of incidents or fatalities. SVMs are efficient in finding and categorizing patterns in the data, and they help locate hotspots and forecast the trends of the pandemic. Overall, in attempts to prevent and mitigate the COVID-19 pandemic, policymakers and public health professionals can benefit from the combination of data visualization and machine learning tools.

References

[1] Sohrabi, C., Alsafi, Z., O'neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., & Agha, R. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery*, *76*, 71–76.

[2] Mohanty, S. K., Satapathy, A., Naidu, M. M., Mukhopadhyay, S., Sharma, S., Barton, L. M., Stroberg, E., Duval, E. J., Pradhan, D., Tzankov, A., & Parwani, A. V. (2020). Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) and coronavirus disease 19 (COVID-19) – anatomic pathology perspective on current knowledge. *Diagnostic Pathology*, *15*(1), 103. https://doi.org/10.1186/s13000-020-01017-8

[3] Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., MacDonald, B., Beltekian, D., & Roser, M. (2020). Coronavirus pandemic

(COVID-19). Published online at OurWorldInData. Org. *Edited By*, *2021*.

[4] McKeever, A. (2021). COVID-19 surpasses 1918 flu as deadliest pandemic in US history. *National Geographic*.

[5] Stobbe, M. (2020). US deaths in 2020 top 3 million, by far most ever counted. *Associated Press News*.

[6] Bosman, J., Kasakove, S., & Victor, D. (2021). US life expectancy plunged in 2020, especially for Black and Hispanic Americans. *The New York Times*, *21*.

[7] Andrasfay, T., & Goldman, N. (2021). Reductions in US life expectancy from COVID-19 by Race and Ethnicity: Is 2021 a repetition of 2020?. *Medrxiv*.

[8] Shapiro, E., Pereira, I., & Deliso, M. (2021). COVID-19 live updates: More Americans died of COVID this year than all of 2020. *ABC News*.

[9] Business Cycle Dating Committee. (2020). Determination of the February 2020 Peak in US Economic Activity,[Press release].

[10] Bright, M. M., Komadina, S., Korntheuer, A., Arslan, A., Bieck, S., Tehrani, K., Goethe, T., Tapee, J., Caldwell, N., & Keding, P. (2020). *Advanced High-Lift Leading-Edge (AHLLE) Technologies for Laminar Wings*. https://ntrs.nasa.gov/citations/20210011080

[11] West, D. M. (2010). *Brain gain: Rethinking US immigration policy*. Brookings Institution Press.

[12] US Bureau of Labor Statistics. (1939). All employees, total nonfarm (PAYEMS) | FRED | St. Louis Fed. stlouisfed.org. *FRED, Federal Reserve Bank of St. Louis*. Retrieved June 4, 2019.

[13] Arfan A., (2021). Visualization and Prediction of COVID-19 future outbreak by using machine learning. *International Journal of Information Technology and Computer Science, 6*(2), pp17-34.

[14] Arıtürk, M., Yavuz, S., & Allahviranloo, T. (2020). Artificial Intelligence and Autonomous Car. *Mathematical Methods in Interdisciplinary Sciences*, 391–412.

[15] Engledowl, C., & Weiland, T. (2021). Data (Mis) representation and COVID-19: Leveraging misleading data visualizations for developing statistical literacy across grades 6–16. *Journal of Statistics and Data Science Education*, *29*(2), 160–164.

[16] Awodiji, T. O. (2023). Future maintenance and service innovation using industrial big data analytics in the United States. *Computer Engineering and Intelligent System, 4*(1), 26

[17] Yuen, M. C., Cheng, W. F., & Wong, H. H. (2022). Data Visualization and Analysis with Machine Learning for the USA's COVID-19 Prediction. In *Fuzzy Systems and Data Mining VIII* (pp. 181–190). IOS Press.

[18] Arogyaswamy, B. (2020). Big tech and societal sustainability: An ethical framework. *AI & Society*, *35*(4), 829–840.

[19] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144.

[20] Gonsalves, C., & Nitnaware, P. (2021). Prediction of Covid-19 Cases Using Machine Learning. *Journal of Emerging Technologies and Innovative Research, 10*(8), a518-a530.

[21] Gonsalves, C., & Nitnaware, P., (2021). Prediction of Covid-19 Cases Using Machine Learning. *Journal of Emerging Technologies and Innovative Research, 10*(8), a518-a530.

[22] Gonsalves, C., & Nitnaware, P., (2021). Prediction of Covid-19 Cases Using Machine Learning. *Journal of Emerging Technologies and Innovative Research, 10*(8), a518-a530.

[23] Jiang, X., Nie, X., Guo, X., Song, C., & Chen, J. G. (2020). Recent advances in carbon dioxide hydrogenation to methanol via heterogeneous catalysis. *Chemical reviews*, *120*(15), 7984-8034.

[24] Alahmari, S. A. (2019). Using Machine Learning ARIMA to Predict the Price of Cryptocurrencies. *ISeCure*, *11*(3).

[25] Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review*, *53*, 5929–5955.

[26] Sharmila, S. L., Dharuman, C., & Venkatesan, P. (2017). Disease classification using machine learning algorithms-a comparative study. *International Journal of Pure and Applied Mathematics*, *114*(6), 1-10. *Mathematics. 114(*6), pp1–10.

[27] Tiwari, U. K., & Khan, R. (2020). Role of machine learning to predict the outbreak of covid-19 in India. *Journal of Xi'an University of Architecture & Technology*, *12*, 2663–2669.

[28] Theodoros, P., & Massimiliano, P. (2022). Workshop on support vector machines: Theory and applications. *Center for Biological and Computational Learning, and*

Artificial Intelligence Laboratory, MIT E25-201.

[29] Georgevici, A. I., & Terblanche, M. (2019). Neural networks and deep learning: A brief introduction. *Intensive Care Med* **45**, 712–714. https://doi.org/10.1007/s00134-019-05537-w

[30] Di Franco, G., & Santurro, M. (2021). Machine learning, artificial neural networks, and social research. *Journal of Qualitative and Quantitative Networks,* **55**, 1007–1025. https://doi.org/10.1007/s11135-020-01037-y

[31] Park, E., Chang, H.J., & Nam, H.S. (2018). A Bayesian network model for predicting post-stroke outcomes with available risk factors. *Frontiers in Neurology*, 9, 2-14. https://doi.org/10.3389/fneur.2018.00699

[32] Mitchell, T. M., & Learning, M. (1997). Mcgraw-hill science. *Engineering/Math*, *1*, 27.

[33] Adelson-Velskiy, G. M., Arlazarov, V. L., Donskoy, M. V., Elman, T. S., & Hinton, G. (1997). Acharya, A., see Chakrabarti, PP. Adam, A. and J.-P Laurent, LAURA, a system to debug stu-dent programs Adcock, J., see Char&k, E. Addanki, S., R. Cremonini and JS Penberthy, Graphs of mod. *Artificial Intelligence*, *96*, 229-299.

[34] Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 249–256.

[35] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision*, 1026–1034.

[36] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *ArXiv Preprint ArXiv:1412.6980*.

[37] Hassan, A. H. M., Qasem, A. A. M., Abdalla, W. F. M., & Elhassan, O. H. (2021). Visualization & prediction of COVID-19 future outbreak by using machine learning. *Int. J. Inf. Technol. Comput. Sci*, *13*(3), 16–32.

[38] Huang, H., & Abdel-Aty, M. (2010). Multilevel data and Bayesian analysis in traffic safety. *Accident Analysis & Prevention*, *42*(6), 1556–1565. doi: 10.1016/j.aap.2010.03.013.

[39] Dey, S. K., Rahman, M. M., Siddiqi, U. R., Howlader, A., Tushar, M. A., & Qazi, A. (2022). Global landscape of COVID-19 vaccination progress: insight from an exploratory data analysis. *Human vaccines & immunotherapeutics*, *18*(1), 2025009.

[40] Cahill, G., Kutac, C., & Rider, N. L. (2021). Visualizing and assessing US county-level COVID19 vulnerability. *American Journal of Infection Control*, *49*(6), 678-684.

[41] Awodiji, T.O. (2021). Database security in a dynamic IT world. *7th International Conference on Software Engineering and Applications (SOFEA),* 152-161. DOI: 10.5121/csit.2021.111613

[42] Awodiji, T.O. (2021). Database security in a dynamic IT world. *7th International Conference on Software Engineering and Applications (SOFEA)*, 152-161. DOI: 10.5121/csit.2021.111613