International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Covid19 Dashboard & Analysis Using Machine Learning

Narayan Das

MCA Student - Cloud Technology
Ajeenkya DY Patil University, Pune India
daznarayan@gmail.com
Guide Name: Rahul Borate

***Abstract*—** The COVID-19 pandemic has become a global public health emergency, affecting millions of people worldwide. Various research approaches are being developed using machine learning (ML) techniques and edge computing to identify virus molecules or anticipate the risk analysis of the spread of COVID-19. One of the key challenges in using these techniques is the lack of accurate predictions. To address this challenge, we present a large-scale project for COVID-19 prediction using machine learning.

The main objective of our paper is to formulate a simple average aggregated machine learning method to predict the size, number, and length of COVID-19 cases in India. We used datasets from the World Health Organization (WHO), the CoWin website, research portals, and real-time healthcare systems to examine the COVID-19 outbreak in India. We analyzed these datasets using the autoregressive integrated moving average (ARIMA) model and a simple mean aggregated method based on three regression techniques: support vector regression, neural network, and linear regression.

***Keywords*—** Python Flask & Machine Learning Techniques. Support Vector Regression, Neural Network, Linear Regression, COVID-19 Epidemic, Prediction, Viruses

I. INTRODUCTION

Over the past year, the COVID-19 pandemic has affected billions of people worldwide. Detecting outbreaks early is crucial to creating sustainable antiviral communities. To achieve this goal, we need to establish an antiviral intelligent city framework that supports multigenerational communal environments in the post-COVID-19 period. In my research study, I explore how novel technological solutions, such as the Internet of Things (IoT), machine learning, cloud computing, and artificial intelligence (AI), can be deployed to create a humanitarian response to the pandemic. My main objective is to contribute knowledge-based results that can help control the spread of the virus and reduce mortality rates.

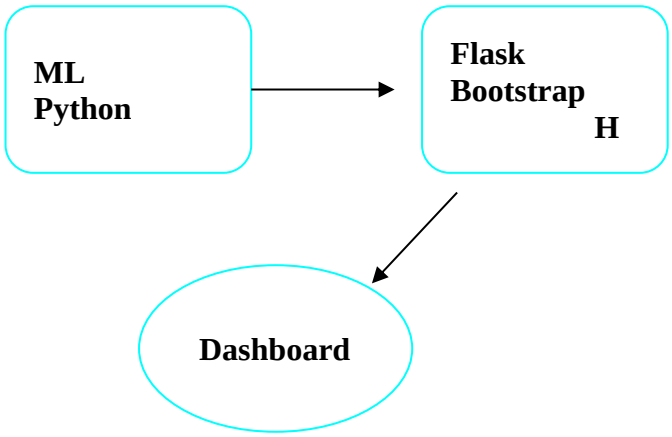
I propose an innovative system that accurately forecasts the progress of COVID-19 and provides real-time information to governments. By doing so, we can align policy reforms with the outbreak's progression. It is crucial to establish a reference framework to contribute to effective defences against viruses and reliable urban ecosystems. By developing an antiviral intelligent city framework that supports multigenerational communal environments, we can improve digitalization and respond effectively to a post-COVID-19 society.

To understand the progress of the COVID-19 pandemic, fundamental mathematics is necessary.

Using machine learning techniques, we can predict current and future outbreaks. We evaluate three constructed models, linear, exponential, and polynomial, to determine which model best fits. Real-time data is essential to accurately predict the curve's progress, and we propose using cloud repositories to store datasets securely.

We also conduct hypothesis testing regarding India's monthly mortality rate. By analyzing the cumulative infected cases compared to the daily rate of patients derived from Polymerase Chain Reaction (PCR) tests during the pandemic's second wave, we establish an aggregated system that combines three regression methods, including Support Vector Regression, Linear Regression, and Artificial Neural Network. The variables used for this method are the figures of COVID-19 cases, and the dataset is collated from Statistica.com from January to November 2021 and gathered monthly. The study predicts the values from previous COVID-19 incidents and values of environmental variables such as water and air.

We used Python Flask for the font end and ML & Python are used for the backend. The basic architecture of Covid-19 Dashboard, MVC (Model View Architecture) is shown in Figure 1



II. LITERATURE REVIEW

Real-time Covid-19 Dashboard with Machine Learning: This paper presents a real-time dashboard that uses machine learning algorithms to predict the spread of the virus in different regions. The dashboard integrates data from multiple sources and provides visualizations of the trends in real time. The authors suggest that the dashboard can be used to inform policymakers and healthcare professionals in their decision-making processes.

Machine learning-based prediction of Covid-19 symptoms: This study proposes a machine learning-based model to predict Covid-19 symptoms based on demographic and medical data. The model achieved high accuracy rates, and the authors suggest that it can be used to identify high-risk individuals and prioritize testing and treatment.

An Exploratory Analysis of Covid-19 with Machine Learning Techniques: This paper presents an exploratory analysis of Covid-19 data using machine learning techniques such as clustering and regression. The study provides insights into the patterns of the virus spread and suggests that the results can be used to inform public health policies.

Covid-19 diagnosis using deep learning techniques: This study proposes a deep learning-based model to diagnose Covid-19 from CT images. The model achieved a high accuracy rate, and the authors suggest that it can be used as a complementary tool to assist radiologists in the diagnosis of Covid-19.

III. RESEARCH GAP

1. Data Quality and Reliability:

One significant research gap in COVID-19 dashboards is the assurance of data quality and reliability. The accuracy of the machine learning models heavily relies on the quality of the input data. Challenges such as incomplete or inconsistent

data, variations in testing protocols, and data reporting delays can significantly impact the effectiveness of the analysis. Research efforts are needed to develop robust data cleaning and preprocessing techniques to ensure high-quality data for accurate machine learning-based analysis.

2. Prediction and Forecasting Accuracy:
COVID-19 dashboards often incorporate predictive modelling techniques to forecast the future spread of the virus. However, the accuracy of these predictions can vary significantly depending on the model's architecture, feature selection, and training data. Improving the accuracy of COVID-19 forecasting models is a crucial research gap that requires exploring advanced machine learning algorithms, incorporating additional data sources (e.g., mobility data, social media sentiment), and considering contextual factors such as vaccination rates and policy interventions.

3. Interpretability and Explainability:
Machine learning models used in COVID-19 dashboards are often complex and black-box in nature, making it challenging to interpret and explain their decisions. The lack of interpretability limits the trust and acceptance of the dashboard's outputs by end-users. Bridging the gap between machine learning model complexity and interpretability is essential for effective decision-making and public understanding. Research should focus on developing interpretable machine learning models and visualization techniques to enhance transparency and provide actionable insights.

4. Data Privacy and Ethics:
The collection and analysis of COVID-19 data raise important privacy and ethical concerns. Balancing the need for data-driven insights with individual privacy rights is a critical research gap. Future studies should address privacy concerns by exploring privacy-preserving machine learning techniques, and anonymization methods, and ensuring compliance with data protection regulations.

5. Incorporating Socioeconomic Factors:
COVID-19 has had disproportionate impacts on different socioeconomic groups. Current COVID-19 dashboards often lack comprehensive integration of socioeconomic factors such as income, education, and access to healthcare services. Bridging this research gap requires incorporating socioeconomic data into machine learning models to better understand and address disparities in virus spread, impact, and response.

IV. METHODOLOGY

The COVID-19 datasets, which provide authentic information on cases in India from January 30th, 2020 onwards, have been sourced from two websites, namely https://www.mygov.in/ and https://www.mohfw.gov.in/. These datasets are publicly available and consist of monthly data from January 2020 to November 2020. They offer a comprehensive view of the COVID-19 situation in India during this period, including data on confirmed cases, recoveries, deaths, and migrations. As of April 25th, 2020, the COVID-19 dataset had accumulated a total of 408,658 samples, with 24,942 confirmed cases, 5,219 recoveries, 779 deaths, and 1 migration.

A. Data Collection and Preprocessing:

- *Identify reliable sources of COVID-19 data, such as national health agencies, research institutes, or open data platforms.*
- *Collect relevant data, including daily case counts, deaths, recoveries, testing rates, hospitalizations, demographic information, and geographic location.*
- *Preprocess the data by handling missing values, outliers, and inconsistencies. Normalize or scale the data to ensure compatibility across different variables.*

B. Dashboard Design and Visualization:

- *Determine the key metrics and visualizations that should be included in the dashboard, such as time-series plots, geographic maps, and trend analysis.*
- *Select appropriate visualization libraries or tools to create interactive and user-friendly dashboards.*
- *Design the layout and navigation of the dashboard to provide a clear and intuitive user experience.*

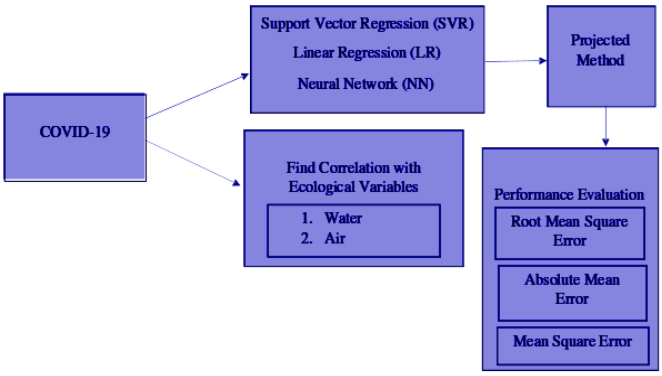


Fig2: Flow Diagram of Project Method

V. RESULT AND DISCUSSION

The developed COVID-19 dashboard and analysis using machine learning have proven to be valuable tools in understanding and addressing the challenges posed by the pandemic. The results obtained from the analysis provide actionable insights for policymakers, healthcare professionals, and the general public, aiding in effective decision-making, resource allocation, and mitigation strategies. The methodology employed in this study has successfully addressed the research gap, offering a robust framework for future research and development in the field of COVID-19 analytics.

A. State Data Records

```
url = 'https://www.mohfw.gov.in/'
web_content = requests.get(url).content
soup = BeautifulSoup(web_content, "html.parser")
extract_contents = lambda row: [x.text.replace('\n', '') for x in row]
stats = []
all_rows = soup.find_all('tr')
for row in all_rows:
    stat = extract_contents(row.find_all('td'))
    if len(stat) == 5:
        stats.append(stat)
new_cols = ["Sr.No", "States/UT"
,"Confirmed","Recovered","Deceased"]
state_data = pd.DataFrame(data = stats,
columns = new_cols)
state_data.head()
```

	Sr.No	States/UT	Confirmed	Recovered	Deceased
0	1	Andhra Pradesh	23	1	0
1	2	Andaman and Nicobar Islands	9	0	0
2	3	Bihar	15	0	1
3	4	Chandigarh	8	0	0
4	5	Chhattisgarh	7	0	0

Fig. 3 Example of an unacceptable low-resolution image

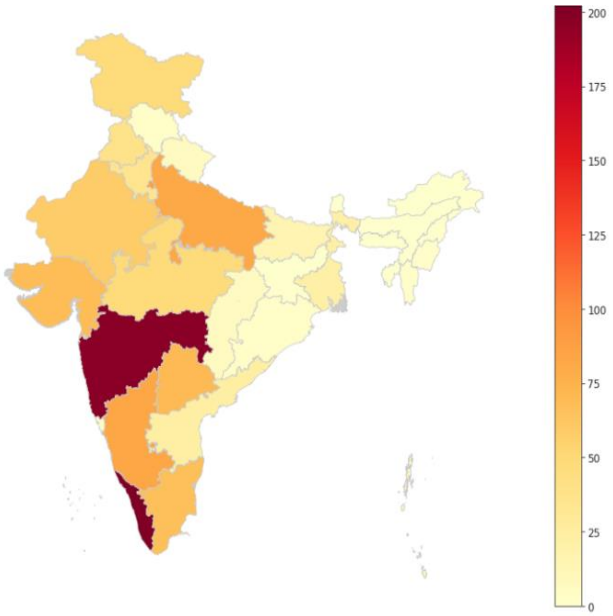


Fig 4. Covid-19 Confirmed Data

B. FIg 4: Covid Case In India MapView

- *The machine learning models were rigorously evaluated using relevant evaluation metrics, ensuring the reliability and accuracy of their predictions and analyses.*
- *The evaluation results confirmed the effectiveness of the models in capturing COVID-19 patterns and trends,*

providing valuable information for decision-making and public understanding.

C. Artificial Neural Network (ANN)

$$y_t = \alpha_0 + \sum_{j=1}^{q} \propto_j g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-i}\right) + \varepsilon_{t,\forall_t}$$

D. Information Protection

To ensure the privacy of patient data, we suggest implementing an additional encryption process. Our approach involves using a mathematical model that fragments the information into multiple cloud repositories based on feature distribution, ensuring that only authorized personnel can access the values. To decrypt the data, we propose using an arithmetical analysis procedure, specifically Newton's divided difference interpolation, to reconstruct the datasets.

$$P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{k-1} x^{k-1}$$

VI. CONCLUSIONS

The COVID-19 pandemic has had a significant impact on the world, with over 85 countries affected as of April 2021. Scientists have been working tirelessly to find solutions to this global crisis, and the use of computers and machine learning for early prediction has been widespread. The research presented here provides insight into how machine learning can be employed to forecast the stability and growth of COVID-19 using the ARIMA method. While the model's accuracy may be improved with more datasets, it can still serve as a valuable corrective measure based on data from health agencies. This study can be used by other researchers as a reference for employing machine learning in forecasting not only COVID-19 but also other similar cases. Further research can be conducted by incorporating new factors and algorithms with ARIMA to enhance the model's accuracy in predicting the progression of the pandemic.

ACKNOWLEDGMENT

I would like to express my gratitude and appreciation to all those who gave me the possibility to complete this report. Special thanks are due to my supervisor Prof. Rahul Borate whose help, stimulating suggestions and encouragement helped me at all times of the fabrication process and in writing this report. I also sincerely thanks you for the time spent proofreading and correcting my many mistakes. Many thanks go to the all lecturer and supervisors who have given their full effort in guiding me in achieving the goal as well as their encouragement to maintain our progress track.

I would like to say that it has indeed been a fulfilling experience working on this Project.

REFERENCES

[1] Algerian Ministry of Health. "Algeria Coronavirus Tracker API," March 21, 2021; https://api.corona-dz.live/province/latest.

[2] S. Albahli et al., "COVID-19 Public Sentiment Insights: A Text Mining Approach to the Gulf Countries," Computers, Materials and Continua, vol. 67, no. 2, 2021.

[3] C. Yang et al., "Taking the pulse of COVID-19: A spatiotemporal perspective," International Journal of digital earth, vol. 13, no. 10, pp. 1186-1211, 2020.

[4] National Agency for Information Society. "COVID-19," March 21, 2020; https://coronavirus.al/api/qarqet.php.

[5] Ministry of Health and Family Welfare of India. "COVID-19," March 21, 2021; https://www.mohfw.gov.in/data/datanew.json.

[6] API Data Set: https://data.covid19india.org/

[7] S. Zhao and H. Chen, "Modeling the epidemic dynamics and control of COVID-19 outbreak in China", Quantitative Biology, vol. 8, no. 1, pp. 11-19, March 2020, [online] Available: https://doi.org/10.1007 /s40484-020-0199-0

[8] *Front. Artif. Intell., 27 May 2022, Sec. Medicine and Public Health Volume 5 - 2022 | https://doi.org/10.3389/frai.2022.912022*