© May 2024 | IJIRT | Volume 10 Issue 12 | ISSN: 2349-6002

COGNIZANCE: Analysis and Prediction of COVID-19 using ARIMA-Random Forest Classifier

AAKRITI LUTHRA[1], PRABHPREET KAUR[2], PARMINDER KAUR[3] AMANDEEP KAUR[4]
[1, 2] Computer Engineering And Technology, Guru Nanak Dev University, Amritsar.
[3] Computer Science, Guru Nanak Dev University, Amritsar.
[4] Computer Engineering And Technology , Guru Nanak Dev University, Amritsar.

Abstract*— The COVID-19 pandemic has unleashed its merciless grip on the world, rapidly escalating with each passing day. In the midst of this crisis, the need for accessible and readily available services has become paramount. The coronavirus outbreak has cast a long, dark shadow over global health, resulting in a devastating loss of life and overwhelming healthcare systems. The World Health Organization's declaration of COVID-19 as a pandemic has thrown every nation into turmoil, straining healthcare infrastructures and prompting often sluggish responses. On a daily basis, the global tally of COVID-19 cases continues to climb, necessitating innovative solutions. Our system, built on Django and utilizing HTML (Hypertext Markup Language), CSS (Cascading Style Sheets), Bootstrap, Python, and Machine Learning Algorithms, has accurately predicted COVID-19 cases using the Random Forest Classifier and ARIMA (Autoregressive integrated moving average) model. System uses techniques that deliver data visualization and analysis Its purpose is to empower users with an array of tools, including locating nearby COVID-19 treatment facilities, accessing in-depth information on medical professionals treating the virus, symptom and risk factor checks, and real-time pandemic updates. The system also provides essential data on vaccinations, live cases, safety measures, and predictive analyses for global and country-specific COVID-19 cases. Users are encouraged to share their experiences through the UserStory column and can engage with administrators for support and feedback. This project additionally lists volunteers providing food to patients and their loved ones and offers valuable information about COVID-19, aiming to enhance user awareness during these trying times.

***Index Terms*— Machine Learning, Covid-19, Statistical Analysis, Prediction, ARIMA, Random Forest.**

I. INTRODUCTION

Severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) are two highly transmissible and pathogenic viruses that emerged in humans at the beginning of the 21st century.[12] The emergence of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Wuhan, China in December 2019 marked the onset of the devastating coronavirus disease 2019 (COVID-19) pandemic.[16] This global crisis has affected millions, with infections and fatalities on a disturbing rise. The vast coronavirus family, including common cold-causing strains, is known for its ability to induce a range of illnesses, from mild to severe. COVID-19 spreads through respiratory droplets, fecal-oral transmission, and direct contact, with an incubation period of two to fourteen days. As there's no specific antiviral therapy, prevention remains the primary approach.

In this scenario, understanding and forecasting the disease's spread is vital for effective decision-making. To combat the growing pandemic on urgent basis, there is need to design efective solutions using new techniques that could exploit recent technology, such as machine learning, deep learning, big data, artifcial intelligence, for identifcation and tracking of COVID-19 cases in near real-time.[13] With its visualization and prediction capabilities, machine learning is globally employed to study COVID-19 transmission. A key research focus is the use of machine learning to analyze, predict, and visualize the virus's global and country-wise spread over time, considering confirmed cases, recoveries, and fatalities.

Introducing 'COGNIZANCE,' a platform aiding users in finding nearby hospitals, COVID-19 treating doctors, symptoms & precautions, tiffin services, blogs, vaccination updates, blogs, and the latest news. It aims to bridge the gap between reported COVID-19 cases and actual incidence. The website also offers insights into COVID-19 vaccines and provides real-time case data in India and worldwide. Users can access live COVID-19 news, global and country-specific case analysis, and a predictive tool to estimate future cases. 'COGNIZANCE' prides itself on user-friendliness and ease of use, contributing to informed decision-making during the pandemic.

1.1 WHAT IS PANDEMIC

A pandemic is the global spread of a novel and fatal disease. Respiratory infections caused by a novel infectious agent or coronavirus are the most susceptible to developing into a global pandemic. A pandemic occurs when a disease spreads across multiple countries or continents, affecting more people and taking more lives. COVID-19 was classified as a pandemic by the World Health Organization (WHO) when it became apparent that the disease was severe and rapidly spreading across a large area.

The pandemic alert system of the World Health Organization (WHO) varies from Phase 1 (low-risk) to Phase 6 (full pandemic):

Phase 1- There are no known human infections caused by a virus in animals [hyp 1].

Phase 2- A human infection caused by an animal virus [hyp 1].

Phase 3- There are always a few cases of illness in humans. If the illness is spreading from human to human, it's not broad enough to cause community-level outbreaks [hyp 1].

Phase 4- With confirmed outbreaks at the Community level, the disease is spreading from person to person [hyp 1].

Phase 5- In more than one country of the WHO regions, the disease is spreading between people [hyp 1].

Phase 6- There has been an outbreak of Community level disease in at least one more country, in a different region from Phase 5 [hyp 1].

A pandemic isn't equal to an epidemic. In an epidemic, there are more cases of illness than would normally occur in a given community or region, but the disease does not spread.

The origin of the pandemic flu virus is usually animal influenza viruses and it does not resemble the seasonal influenza virus, if in fact they are already vaccinated with or have received seasonal influenza vaccine. The outbreak of COVID-19 is illustrated in *Figure* 1.

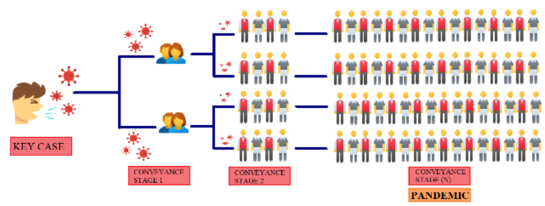

Figure 1: Spreading of COVID-19 outbreak

1.2 TECHNOLOGY USED

1.2.1 HTML

HTML stands for HyperText Markup Language [hyp 2]. It is used to design web pages using a markup language [hyp 2]. HTML is a combination of Hypertext and Markup language [hyp 2]. Hypertext defines the link between web pages. A markup language is used to define the text document within the tag which defines the structure of web pages [hyp 2]. A set of elements called tags are employed to generate the HTML code. The HTML components give the browser the ability to display content. The web page's architecture, which is the first thing a user would interact with, is defined by the HTML code. HTML contains numerous elements that we refer to as tags, such as heading tags (h1 to h6), paragraph tags (p>paragraph/p>), image tags, anchor tags, etc.

The HyperText Markup Language or HTML is the standard markup language for documents designed to be displayed in a web browser [hyp 3]. It defines the meaning and structure of web content [hyp 3]. It is often assisted by technologies such as Cascading Style Sheets (CSS) and scripting languages such as JavaScript [hyp 3].

1.2.2 CSS

The CSS acronym stands for Cascading Style Sheets. Cascading Style Sheets, sometimes known as CSS, is an easy-to-use design language that was developed to make the process of building appealing web pages simpler. CSS controls a web page's style and look. Using CSS, it is possible to change the font shade, style of font, paragraph spacing, size of column, structure, background images or colors, layout designs, deviations in display for different screen sizes and devices, and a wide range of other effects.

Although CSS is relatively easy to acquire and comprehend, it has a huge impact regarding how an HTML text looks. When producing markup, CSS is most typically used in conjunction with HTML or XHTML.

1.2.3 BOOTSTRAP

In recent years, Bootstrap has become the most popular front end framework. It is a sleek, intuitive, and powerful mobile-first front-end framework designed to make web development faster and easier. It makes use of HTML, JavaScript, and CSS. Bootstrap is a framework that allows you to easily develop web projects. Bootstrap is broken down into categories such as Bootstrap CSS, Bootstrap Layout Components, Bootstrap Basic Structure, and Bootstrap Plugins. Bootstrap is a free front-end framework for faster and easier web development and includes HTML and CSS-based design templates for typography, forms, buttons, tables, navigation, modals, image carousels and others, as well as optional JavaScript plugins [hyp 4].

1.2.4 PYTHON

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages [hyp 5]. Another language that's on the rise is Python. It's got a huge library of modules that are useful for integrating complicated solutions from existing components. A large number of individuals are contributing to the Python Open-Source Project. It's a platform-independent, scriptable language that has full access to operating system APIs. This enables user applications to be integrated in such a way that they can create powerful, highly focused applications with ease.

Some of the primary benefits of learning Python include:

Python is Interpreted Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP [hyp 5].

Python is Object-Oriented Python supports Object-Oriented style or technique of programming that encapsulates code within objects [hyp 5].

Python is Interactive You can actually sit at a Python prompt and interact with the interpreter directly to write your programs [hyp 5].

Python is a Beginner's Language Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games [hyp 5].

1.2.5 MACHINE LEARNING (ML)

Machine learning (ML) is a type of artificial intelligence (AI) focused on building computer systems that learn from data [hyp 6]. The broad range of techniques ML encompasses enables software applications to improve their performance over time [hyp 6]. Machine learning algorithms are trained to find relationships and patterns in data [hyp 6]. They use historical data as input to make predictions, classify information, and cluster data points [hyp 6].Machine learning is an application of AI that enables systems to learn and improve from experience without being explicitly programmed [hyp 7]. Machine learning focuses on developing computer programs that can access data and use it to learn for themselves.[hyp 7]

1.2.6 DJANGO

Django is a high-level Python web framework that encourages rapid development and clean, pragmatic design [hyp 8]. Django uses the MVT architecture, that is to say Models, Views and Templates. When we use the Django framework, our project is split into a MVT structure where Models keep our database code and Templates store HTML, CSS, Javascript Code and Views to establish connections with models and Templates. Django is an advanced Python-based web framework that promotes fast development and pure, pragmatic design. With a low amount of code, Django helps to speed up the development of better web apps. Django comes with a set of design philosophies:

Loosely Coupled − Django aims to make each element of its stack independent of the others. [hyp 8]

Less Coding − Less code so in turn a quick development. [hyp 8]

Don't Repeat Yourself (DRY) − Everything should be developed only in exactly one place instead of repeating it again and again. [hyp 8]
Fast Development − Django's philosophy is to do all it can to facilitate hyper-fast development.
Clean Design − Django strictly maintains a clean design throughout its own code and makes it easy to follow best web-development practices. [hyp 8]

II. LITERATURE SURVEY

A vast amount of research studies were conducted for detection, prevention, and analysis of COVID-19. Multiple machine learning algorithms are used to predict the COVID-19. Machine Learning has been applied to data for visualization and prediction. A comprehensive analysis of various research studies conducted to combat the COVID-19 pandemic using machine learning, deep learning, and artificial intelligence techniques. Each study is summarized with its aim, dataset, tools and algorithms used, performance metrics, merits, and demerits.

Aditra Pradnyana et al. (2021) developed a web-based system to analyze the sentiment of Bali's tourism industry during the COVID-19 pandemic. They utilized data from Twitter based on specific keywords related to Bali tourism. The system was built using the Django framework and Python programming language, and the Naïve Bayes algorithm was employed for sentiment analysis. The study achieved an accuracy of 81.70%, precision of 75%, recall of 75%, and an F-measure of 82.02%. The researchers successfully developed a web application for sentiment analysis of Bali tourism during the pandemic. However, the Naïve Bayes algorithm yielded subpar results, and the sentiments relied heavily on language. Additionally, the data collection was specific to Twitter, limiting the scope to English sentiments, and ongoing keyword monitoring and expansion were required. Pourhomayoun and Shakibi (2021) aimed to develop an AI model to determine the health risk and predict the mortality risk of patients with COVID-19. They employed various machine learning algorithms, including Support Vector Machines (SVM), Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Random Forest, Decision Trees, and Logistic Regression. The study utilized Python for implementation and achieved an accuracy of 89%. The developed model assesses health risks and predicts patient mortality, allowing for the prioritization of patient care based on urgency. It utilizes global hospital data, making it a universal and robust model. However, the model's scalability to other diseases needs further exploration, and additional data may be required. The study also lacks a comprehensive evaluation using various performance metrics. Alves et al. (2021) presented understandable solutions based on machine learning techniques to deal with COVID-19 screening in routine blood tests. They used a dataset from the Albert Einstein hospital in Sao Paulo, Brazil. The study employed the Random Forest algorithm and the Decision Tree Explainer (DTX) for interpretability, implemented in Python. The results showed an accuracy of 0.88, an F1 score of 0.76, sensitivity of 0.66, specificity of 0.91, and an AUROC of 0.86. The Random Forest algorithm delivered optimal results, and the solution was both reproducible and interpretable. However, human specialists may doubt the decisions made by the machine learning model, and the study suffered from limited accuracy and an inadequate data size.

Hussain et al. (2022) aimed to predict and evaluate the healthy and unhealthy status of COVID-19 patients. They collected real-time data from an API designed explicitly for quarantined COVID-19 patients, containing various measured parameters. The study utilized Python, Google Colab, Jupyter Notebook, and several machine learning algorithms such as Random Forest Classifier, Linear Regression, Polynomial Regression, SVM, Decision Trees, and Logistic Regression. The results showed an accuracy of 99.26%, a classification rate of 96.8%, and a confusion matrix score of 97.05%. The study addressed challenges in disease diagnosis, and the diagnosis outcomes can gauge patient severity, aiding outbreak management and patient care. However, the study acknowledged the need for more data and the limited availability of parameters. Villavicencio et al. (2022) developed a new approach to disease detection and symptom analysis without the need for lab tests. They utilized a dataset from Kaggle titled "COVID-19 symptoms and presence." The study employed Python, Google Colab, PyCharm IDE, and Azure web services, along with various machine learning algorithms such as Random Forest, SVM, KNN, ANN, Naïve Bayes, and Decision Trees. The results showed an accuracy of 98.84%, sensitivity of 100%, specificity of 98.79%, and an AUROC of 98.84%. The developed system features a symptom checker, researcher contact, feedback form, and a COVID-19 prediction section with 16 symptoms. It facilitates direct communication with researchers and provides links to the WHO website for updated information. However, the study acknowledged that symptoms vary among individuals, and the system lacks a helpline, healthcare service information, and user record-keeping functionality. Kumar Singh et al. (2022) aimed to automate the daily operations and functionality of hospitals, such as fixing appointments, adding doctor and patient details, and providing government-approved guidelines. The study utilized Python, HTML5, CSS, Django, and SQLite3. The developed system covers core multi-specialty hospital functions, enhances patient safety, confidentiality, efficiency, cost control, and symptom management, and aids in protocol maintenance. However, the system lacks online doctor chat and appointment features, and it does not provide information on nearby hospital facilities, testing, or vaccination booking. Miralles-Pechuán et al. (2023) conducted experiments to predict COVID-19 cases for the 50 countries with the most cases during 2020 and compared the performance of various machine learning algorithms. The study used the COVID-19 coronavirus data daily (14/12/20) available in the European open data portal by the European Centre for disease prevention and Control. The study employed Python and various algorithms such as SVM, Random Forest, Gradient Boosting, Linear Regression, Ridge Regression, Decision Trees, Bayesian Ridge Regression, Passive Aggressive algorithm (PA), Adaptive Random Forest (ARF), Adaptive Windowing (ADWIN), ARIMA, and Hoeffding Tree (HT). The results showed a Euclidean distance of 74.21% and a Dynamic Time Warping of 74.89%. The study provided public code for research exploration and applied an Incremental Learning Model (ILM) for COVID-19 prediction. Time series similarity measures improved accuracy. However, the study lacked parameters for virus evolution scenarios, was limited to 50 countries, and the LSTM model performed poorly compared to static methods. Additionally, training with all 50 countries yielded slightly worse results than training with just one country.

Raman et al. (2023) aimed to develop a machine learning model to predict COVID-19 severity at the time of hospital admission based on single-institution data. The study employed Python, the Random Forest Classifier, and K-Nearest Neighbors (KNN) algorithms. The results showed a sensitivity of 0.72 – 0.75, specificity of 0.78 – 0.75, and an AUC of 0.82-0.81. The developed model categorizes scores into user-friendly risk levels. However, the model's performance could be enhanced by including image data such as chest X-rays or CT scans. Paul et al. (2023) assessed the performance and investigated different machine learning, deep learning, and combinations of various ML, DL, and AI approaches that have been employed in recent studies with diverse data formats to combat the problems that have arisen due to the COVID-19 pandemic. The study aimed to outline prior research work and how these works have been used to fight COVID-19. The study utilized Python and various machine learning, deep learning, and artificial intelligence techniques. The study raised public pandemic awareness, utilized ML, DL, and AI-based classification and screening techniques for improved COVID-19 detection, and enhanced COVID-19 management and response efficiency. However, the study highlighted the need for expertise in bioinformatics and relevant domains for successful implementation of ML and DL in COVID-19 studies, and the importance of addressing constraints to build a reliable end-to-end diagnosis solution. Boateng et al. (2023) aimed to predict the mortality rate based on explanatory variables across the five districts of Limpopo Province in South Africa. The study utilized a dataset from the Limpopo department of health and employed Python along with various machine learning algorithms such as Logistic Regression, Random Forest, Support Vector Machines (SVM), Decision Trees, and Random Over-sampling Examples (ROSE). The results showed an AUC of 0.87 for the Random Forest model, with sensitivities of 52% (Random Forest), 51% (Logistic Regression), and 51% (SVM), and a recall of 40% (Decision Tree). The study concluded that the Random Forest model is the most suitable for predicting the COVID-19 mortality rate, as it excels at identifying positive cases compared to other algorithms. However, the study acknowledged that the dataset may not represent the entire population of Limpopo Province, important mortality predictors might be missing, and differences in treatment and care across facilities were not considered.In summary, the provided dataset offers a comprehensive overview of various research studies that employed machine learning, deep learning, and artificial intelligence techniques to combat the COVID-19 pandemic. As discussed in

Table *1*.

Table 1: A summary of different studies related to COVID-19.

REF	AUTHOR, YEAR	AIM, DATASET	TOOLS, ALGORITHM	PARAMETERS	MERITS	DEMERITS
[1]	(Aditra Pradnyana et al., 2021)	-develop web based system to analyze Bali Tourism sentiment -Data from twitter (based on keywords)	-DJANGO -Python -Naïve Bayes	-Accuracy-81.70% -precision-75% -recall-75% -F measure-82.02%	-Developed web app for Bali tourism sentiment analysis during the pandemic successfully.	-Naïve bayes yields subpar results. -Sentiments rely on language. -Twitter- specific data collection. -Limited to English sentiments -Ongoing keyword monitoring and expansion
[2]	(Pourhomayoun & Shakibi, 2021)	-develop AI model to determine the health risk and predict mortality risk of patients with COVID-19	-Python -SVM -ANN -KNN -Random Forest -Decision Tree - Logistic Regression	-Accuracy-89%	-Assesses health risks and predicts patient mortality. -Prioritizes patient care based on urgency -Utilizes global hospital data for a universal and robust model	-Scalable to other diseases -Needs additional data -Limited use of performance metrics
[3]	(Alves et al., 2021)	-present understandable solutions based on ML technique	-Python -Random Forest -Decision Tree Explainer(DTX)	-Accuracy-0.88 -F1 score-0.76 -Sensitivity-0.66 -Specificity-0.91	-Random forest algorithm delivers optimal results -Solution is both	-Human specialist may may doubt ML decisions -Limited accuracy

		to deal COVID-19 screening in routine blood test -Dataset from hospital Albert Einstein, , Sao Paulo, Brazil.		-AUROC-0.86	reproducible and Interpretable	- Inadequate data size
[4]	(Hussain et al.,2022)	-Predict and evaluate healthy and unhealthy status of COVID-19 patients -Realtime data collected can predict the health status of patients from measured parameters -Dataset from API designed explicitly for COVID-19 quarantined patients	-Python -Google Colab -Jupyter Notebook -Random forest classifier -Linear regression -Polynomial Regression -SVM -Decision Tree -Logistic regression	-Accuracy-99.26% -Classification-96.8% -Confusion Matrix-97.05%	-Addressing challenges in disease diagnosis -Diagnosis outcomes can gauge patient severity, aiding outbreak management and care	-Need more data -Limited parameters available
[5]	(Villavicencio et al., 2022)	-develop new approach in disease detection and analyze symptoms experience	-Python -Google colab -DJANGO -Pycharm IDE -Azure web services -Random Forest	-Accuracy-98.84% -Sensitivity-100% -Specificity-98.79% AUROC-98.84%	-Features symptom checker, researcher contact, and feedback form -Facilitates direct communicatio	-Symptoms vary among individuals -Lacks helpline and healthcare service info -No user record keeping

		d without lab test -Dataset from Kaggle "covid-19 symptoms and presence"	-SVM -KNN -ANN -Naïve bayes -Decision Tree		n with researcher -Links to WHO website for updated information -Includes COVID-19 prediction section having 16 symptoms	
[6]	(Kumar Singh et al., 2022)	-Automate daily operation and functionality of Hospitals like fixing appointment, adding doctor details, adding patient details and govt. approved guidelines	-Python -HTML5 -CSS _DJANGO _SQLITE3		-Covers core multi-specialty hospital functions -Enhances patient safety, confidentiality, efficiency, Cost control and symptom management -Aids protocol maintenance	-lacks online doctor chat and appointments -Doesn't provide nearby hospital facilities, testing, or vaccination booking
[7]	(Miralles-Pechuán et al., 2023)	-Conduct experiments to predict COVID-19 cases for 50 countries with most cases during 2020 and compare the performance of ML Algorithms -COVID-19 coronaviru	-Python -SVM -Random forest -Gradient Boosting -Linear regression -Ridge regression -Decision tree -Bayesian Ridge regression -Passive aggressive algorithm(PA)	-Euclidean distance-74.21% -Dynamic Time warping-74.89%	-Public code for research exploration -Model allows continuous updates without retraining -Applies Incremental learning model (ILM) for COVID-19 prediction -Time series similarity measures improve accuracy	-Lacks parameters for virus evolution scenarios -Limited to 50 countries -Poor performance of LSTM model -LSTM underperformed compared to static methods -Training with all 50 countries yielded slightly worse results than training with just one country

		s data daily (14/12/20) available in the European open data portal by the European Centre for disease prevention and Control	-Adaptive random forest (ARF) -Adaptive windowing (ADWIN) -ARIMA -Hoeffding tree (HT)			
[8]	(Raman et al., 2023)	- To develop machine learning model to predict COVID-19 severity at time of hospital admission based on single-institution data	-Python -Random Forest classifier -KNN	-Sensitivity- 0.72 – 0.75 -Specificity- 0.78 – 0.75 -AUC- 0.82- 0.81	-model categorizes scores into user-friendly risk levels.	-model enhancement possible through the inclusion of image data like chest X-rays or CT scans.
[9]	(Paul et al., 2023)	-Assess the performance and investigated different machine learning, deep learning and combinations of various ML, DL, and AI approaches that have been employed	-Python -Machine learning, deep learning, Artificial intelligence		-Raised public pandemic awareness -Utilized ML,DL and AI based classification and screening techniques for improved COVID-19 detection -Enhances COVID-19 management and response efficiency	-Expertise in bioinformatics and relevant domains essential for ML and DL in COVID-19 studies -Address constraints and build a reliable end-to-end diagnosis solution.

		in recent studies with diverse data formats to combat the problem that have arisen due to COVID-19 pandemic - To outline prior research work and how these works have been used to fight COVID-19				
[10]	(Boateng et al., 2023)	-To predict the mortality rate based on explanatory variables across the five districts of Limpopo Province in South Africa -Dataset from Limpopo department of health	-Python -Logistic regression -Random forest -SVM -Decision Tree -Random over-sampling examples(ROSE)	-AUC-0.87(Random forest) -Sensitivity -52% (random forest) -51% (Logistic regression) -51% (SVM) -Recall - 40%(decision tree)	-Random forest is most suitable model for predicting COVID-19 mortality rate. -Random forest excels at identifying positive cases compared to other algorithms	-Dataset may not represent of the entire population of Limpopo Province -Important mortality predictors might be missing -Differences in treatment and care across facilities were not considered

III. PROPOSED WORK

The research aimed to conduct an analysis and make predictions based on a COVID-19 dataset. The data, sourced from respiratory data, underwent necessary cleaning procedures to ensure data quality. This cleaned data was then employed for model training. The study utilized two models, namely the ARIMA model and the Random Forest classifier. The ARIMA model, generally known as the Box-Jenkins methodology is used for forecasting and analysis in the time series approach.[14] To facilitate the use of the ARIMA model, the data was transformed into a time series format, as depicted in Figure 2. Subsequently, the model's performance was assessed, and further analysis and predictions were conducted on the data.

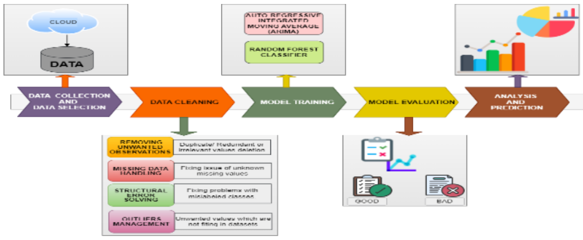

Figure 2: Process Flow of The Research

3.1 DATASET DESCRIPTION

The method used for the Data classification is Random Forest Algorithm and ARIMA model for Machine Learning. Using Random Forest and the ARIMA model, generic machine learning is used to create a diagnosis model for COVID-19 that can identify cases, test results, vaccinations, and human symptoms. The random forest method is a classifier designed to identify the disease using COVID-19 patients' signs and symptoms. ARIMA model is used for data forecasting based on the available data. Every step involved in creating the model, such as loading the dataset, required importing the dependencies. The size of data available is huge and gathering information and getting an interesting pattern out of the cumulated data is a challenging task. [15] The datasets are used for analysis and visualization of available data for better understanding. A readily available Kaggle dataset [11] for data collection is used. In total six datasets are used in the Project. As discussed in Figure 3.

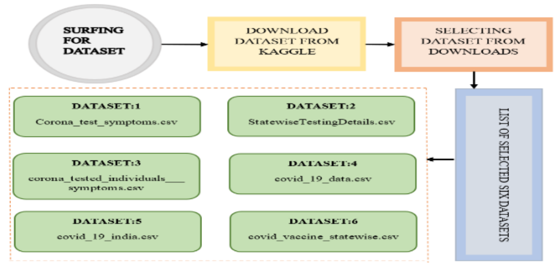

Figure 3: Dataset download and selection layout

The dataset "Corona_test_symptoms.csv" contains 10 attributes and a total of 119 rows. Another dataset, "StatewiseTestingDetails.csv," is composed of 5 attributes and has a substantial 16,336 rows. Similarly, the dataset "corona_tested_individuals_symptoms.csv" consists of 5 attributes and a significantly larger 161,585 rows. In addition, the dataset "covid_19_data.csv" features 8 attributes and an extensive 306,429 rows, while the "covid_19_india.csv" dataset has 9 attributes and a total of 18,110 rows. Lastly, the dataset "covid_vaccine_statewise.csv" encompasses 24 attributes and comprises 7,845 rows. These diverse datasets provide valuable information for various aspects of the COVID-19 pandemic and related research. The descriptions of the dataset's attributes are provided in

Table *2*.

Table 2: Covid-19 predictors and descriptions of the COVID-19 symptoms test dataset.

ATTRIBUTE NAME	DESCRIPTION
TEST_DATE	Date of Test
COUGH	Continuous coughing
FEVER	Temperature is above normal
SORE_THROAT	Experiencing pain, scratchiness, and irritation in the throat
SHORTNESS_OF_BREATH	Experiencing breathing problem such as: shortness of breath, having trouble breathing
HEAD_ACHE	Pain in the head or in a certain part of the head
CORONA_RESULT	The presence of COVID-19
AGE_60_AND_ABOVE	Age of participant above 60 years
GENDER	Gender of the participant: male/ female
STATE	Name of Indian states
TOTALSAMPLES	Total Number of test sample on certain day in certain state

NEGATIVE	Total Number of Negative test sample on certain day in certain state
POSITIVE	Total Number of Negative test sample on certain day in certain state
PROVINCE/STATE	State of a certain country
COUNTRY/REGION	Country of the world
LAST UPDATE	Date of Updation
CONFIRMED	Number of Confirmed cases
DEATHS	Number of death cases
RECOVERED	Number of Recovered cases
CONFIRMEDINDIANNATIONAL	Number of confirmed cases of Indian residents
CONFIRMEDFOREIGNNATIONAL	Number of confirmed cases of Non-Resident or Foreign residents
TOTAL DOSES ADMINISTERED	Total number of dose of vaccine given
SESSIONS	Total number of sessions of vaccin
SITES	Total number sites of vaccine
FIRST DOSE ADMINISTERED	Total number of first dose of vaccine given
SECOND DOSE ADMINISTERED	Total number of second dose of vaccine given
MALE (DOSESADMINISTERED)	Total number of dose of vaccine given to male
FEMALE(DOSESADMINISTERED)	Total number of dose of vaccine given to female
TRANSGENDER(DOSESADMINISTERED)	Total number of dose of vaccine given to transgender
COVAXIN (DOSES ADMINISTERED)	Total number of dose of covaxin given
COVISHIELD (DOSESADMINISTERED)	Total number of dose of covishield given
SPUTNIK V (DOSES ADMINISTERED)	Total number of dose of sputnik v given
AEFI	Total number of adverse events following immunization cases
18-44 YEARS(DOSESADMINISTERED)	Number of dose of vaccine given for age group 18-44
45-60 YEARS(DOSESADMINISTERED)	Number of dose of vaccine given for age group 45-60
60+ YEARS(DOSESADMINISTERED)	Number of dose of vaccine given for age group 60+
18-44YEARS(INDIVIDUALSVACCINATED	Number of individual vaccinated for age group 18-44
45-60YEARS(INDIVIDUALSVACCINATED	Number of individual vaccinated for age group 45-60
60+ YEARS(INDIVIDUALSVACCINATED)	Number of individual vaccinated for age group 60+
MALE(INDIVIDUALS VACCINATED)	Number of individual male vaccinated
FEMALE(INDIVIDUALS VACCINATED)	Number of individual female vaccinated
TRANSGENDER(INDIVIDUALSVACCINATED)	Number of individual transgender vaccinated
TOTAL INDIVIDUALS VACCINATED	Total number of individuals vaccinated

IV. METHODOLOGY

As the COVID-19 pandemic has escalated across the world, researchers have rapidly launched an enormous number of research studies aimed [17] at the visualization and forecasting of data. The data is selected, collected, and downloaded from the databases. Further, data is pre-processed and converted to time-series data. Then, the data is split and the model is selected. On this basis, forecasting and visualized results are obtained. Further, as described in Figure 4.

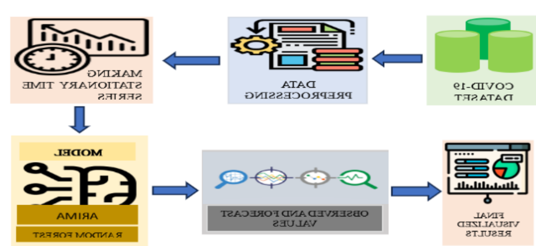

Figure 4: Schematic representation of COVID-19 prediction model with statistical analysis

Amid the global escalation of the COVID-19 pandemic, the need for accurate data visualization and forecasting has spurred an extensive array of research endeavors. Methodology outlines the key steps involved in these research efforts:

4.1 DATA SELECTION

The very first step involves identifying and selecting the most relevant and comprehensive datasets related to COVID-19. Datasets are typically obtained from reputable sources. The project uses data from an open-sourced repository “Kaggle”. The data is freely available for public use and analysis of the pandemic.

4.2 DATA COLLECTION

Once identified, the chosen datasets are systematically collected. The process involves API calls, and manual downloads, ensuring that the data is up-to-date and accurate.

4.3 DATA PRE-PROCESSING

To prepare the collected data for analysis, a series of pre-processing steps are performed. This includes data cleaning, handling missing values, and dealing with outliers. Standardization and normalization techniques are applied to ensure data consistency.

4.4 TIME SERIES CONVERSION

Given the temporal nature of COVID-19 data, it is converted into time-series data. This transformation facilitates the analysis of trends and patterns over time, allowing for more accurate forecasting.

4.5 MODEL SELECTION

Selecting the appropriate model for your analysis depends on the nature of your data and your specific objectives. ARIMA (AutoRegressive Integrated Moving Average) and Random Forest Classifier are two different types of models used for distinct purposes. Selecting the appropriate model for analyzing COVID-19 data, which involves time series forecasting and potentially classification tasks, is a complex undertaking.

4.5.1 ARIMA MODEL

ARIMA is a time series forecasting model that is particularly useful for analyzing and predicting time-dependent data, such as COVID-19 cases over time. Capture trends, seasonality, and autocorrelation in time series data, making it suitable for short- to medium-term forecasting. ARIMA is often used for predicting future values of a continuous variable in a time series, making it useful for applications like financial forecasting, demand forecasting, and epidemiological predictions. ARIMA is applied to predict the number of COVID-19 cases in a specific region over a given time frame [18], helping with resource allocation and public health planning.

ARIMA models are suitable when:

1. Time Series Data: The data is in the form of a time series, with observations recorded at regular intervals (e.g., daily COVID-19 cases, weekly hospitalizations).
2. Stationarity: The time series data exhibits stationarity or can be made stationary through differencing. Stationarity means that the statistical properties (mean, variance, and autocorrelation) of the series remain constant over time.
3. Autocorrelation: The observations in the time series are correlated with their past values. ARIMA models capture this autocorrelation structure, making them useful for forecasting future values based on past observations.
4. Forecasting: The primary objective is to forecast future values of the time series, such as predicting the number of COVID-19 cases or hospitalizations in the upcoming weeks or months.

4.5.2 RANDOM FOREST

Random Forest is a machine learning (ML) model used for classification tasks. It is designed to predict categorical outcomes. Random Forest is known for its robustness, accuracy, and ability to handle both numerical and categorical data. It is also resistant to overfitting and can handle large datasets. A Random Forest Classifier is suitable for classification tasks, such as identifying high-risk areas, demographic groups, and symptoms contributing to the spread of COVID-19. Random Forest is a powerful ensemble learning method that combines multiple decision trees to improve predictive performance and handle complex, non-linear relationships in the data. The Random Forest Classifier is particularly useful for classification tasks, where the goal is to predict a categorical target variable based on a set of independent variables (features).

The Random Forest Classifier is suitable when:
1. Classification Task: The objective is to classify instances into distinct classes or categories. In the context of COVID-19, this could involve classifying patients as high-risk or low-risk based on their symptoms, demographics, and other factors.
2. Multiple Features: The data includes multiple independent variables (features) that can contribute to the prediction or classification task. For COVID-19, these features could include age, gender, pre-existing conditions, symptoms, and laboratory test results.
3. Non-linear Relationships: The relationship between the features and the target variable is complex and non-linear, which is often the case in healthcare and epidemiological data.
4. Robustness: Random Forest models are robust to overfitting, noise, and outliers in the data, making them suitable for real-world datasets that may be imperfect or have missing values.

The system analysis involves a combination of time series forecasting and classification tasks within the same project. Project use each model for its respective purpose.

4.6 FORECASTING AND VISUALIZATION

With the model in place, the data is used to generate forecasts of COVID-19 trends, such as infection rates, mortality, Number of vaccinations, Comparison on the basis of age, Top States and Countries affected, and recovery rates. These forecasts are visualized using charts, graphs, and dashboards to communicate the findings effectively. With the appropriate model in place, such as ARIMA for time series forecasting or Random Forest Classifier for classification tasks, the COVID-19 data can be analyzed to generate valuable insights and forecasts. These forecasts can cover various aspects of the pandemic, including infection rates, mortality, vaccination rollout, age-based comparisons, regional hotspots, and recovery rates. Effective visualization techniques are then employed to communicate these findings clearly and concisely.

Infection Rate Forecasting: By leveraging the time series nature of COVID-19 case data, the ARIMA model can forecast future infection rates at different geographical levels (national, state, or regional). These forecasts can assist public health authorities in anticipating potential surges or declines in cases, enabling proactive measures and resource allocation.

Mortality Projections: COVID-19 mortality data, when analyzed using appropriate models, can provide projections of potential deaths over time. These projections can be further segmented by age groups, underlying health conditions, or other relevant factors, helping healthcare systems prepare for increased demand and allocate resources accordingly.

Vaccination Rollout Monitoring: As vaccination campaigns progress, time series models can forecast vaccination rates and project the timeline for achieving herd immunity. This information is crucial for policymakers to plan and adjust vaccination strategies, ensuring equitable and efficient distribution.

Age-based Comparisons: Random Forest Classifiers can be employed to identify age groups at higher risk of severe COVID-19 outcomes based on various factors such as comorbidities, socioeconomic status, and access to healthcare. These insights can guide targeted interventions and prioritize vaccination efforts for vulnerable populations.

Regional Hotspot Identification: By analyzing COVID-19 data at a granular level (states, counties, or cities), machine learning models can identify emerging hotspots or regions with higher transmission rates. This information can guide the deployment of additional testing resources, contact tracing efforts, and localized containment measures.

Recovery Rate Analysis: Recovery data can be analyzed to understand factors contributing to successful recoveries, such as treatment protocols, patient characteristics, or healthcare system capacities. This information can inform best practices and improve patient outcomes.

Visualization Techniques: To effectively communicate these forecasts and insights, various visualization techniques are employed. Interactive dashboards can display real-time updates on infection rates, mortality projections, and vaccination progress. Charts and graphs can illustrate trends, comparisons, and correlations between different variables. Geographical maps can highlight regional hotspots or vaccination coverage areas.

Effective visualization not only aids in understanding the complex COVID-19 data but also facilitates data-driven decision-making by policymakers, healthcare professionals, and the general public. Clear and concise communication of these forecasts and insights is crucial for raising awareness, guiding resource allocation, and implementing targeted interventions to mitigate the impact of the pandemic.

In summary, methodology encompasses the critical stages of data selection, collection, pre-processing, time-series conversion, model selection, and the subsequent visualization and forecasting of COVID-19 data. It forms the foundation for rigorous research and contributes to our understanding of the pandemic's dynamics and trends.

4.7 INPUTS

Inputs from the administrator and user interfaces shall be taken into account in our system. So there are two sections that can be classified as inputs:

4.7.1 Input by Admin- Input by admin can be in the form of Login, or Managing blogs, Admin will add/edit/delete Blogs. Admin can Change/Recover passwords, Manage Hospitals, Admin will add/edit/delete Hospitals, Manage Doctors, Admin will add/edit/delete Doctors treating COVID-19, Manage Frequently Asked Questions, Admin will add/ edit/ delete FAQs, Manage The Helpline Number, Admin will add/edit/delete Helpline Related Information, Manage Symptoms, Admin will add/ edit/ delete Symptoms of COVID-19 according to the Latest Update, Manage UserStories, Admin will add/edit/ delete UserStories following the publishing protocols, Manage TiffinServices, Admin will add/ edit/ delete TiffinServices, Manage Video's, Admin will add/edit/delete COVID Video's, Manage Contacts, Admin will add/edit/delete Contacts. Admin can Manage Reviews.

4.7.2 Input by User- Input by the user can be of the form Sign up/login. Basic information like user name, email address, and password may be used to register yourself. Users can Change/Recover passwords. The statistics on disease and their effect already made available can be viewed by registered users. By including fields of data needed users are able to build visualizations according to their individual needs. By adding the required fields of data, the user can make predictions. Users can add/ delete UserStory. Users can add reviews/enquiry/suggestions to the admin. The user can send a message to the administrator. Live cases can be checked by users. Users can view their profile and can change their basic information such as Email, Password, phone number, and profile photo.

4.8 CONSTRAINTS/CONDITIONS

To keep the system hustle-free, it will operate under a few conditions. To keep this system running smoothly, there are several conditions which must be met such as: No user shall have access to the Web Application Manager dashboard unless registered. Only registered users can make visualizations according to them. Only a registered user can make predictions based on them. Creating UserStories by registering as a user. The feature of live cases is available only to registered users.

4.9 OUTPUT

The user will be able to make all the available visualizations and predictions listed in our web app. Users will be able to see all the live cases. Admin will see all the added blogs, hospitals, and doctors, vaccinations, FAQs and will be able to add more.

V. EXPERIMENTATION

The subsection shows the results in different tables. A management model has been proposed that has a number of key roles as discussed in Figure 5.

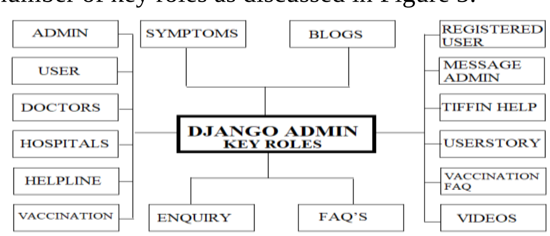

Figure 5: DJANGO Admin Key Roles

Numerous sub-modules specifying their functionalities exist for each of the aforementioned modules:

Multiple tables have been created in the DJANGO database. Figure 6 describes the table names and their schema along with the variable names and their relationship associated with the tables such as Foreign Key(FK).

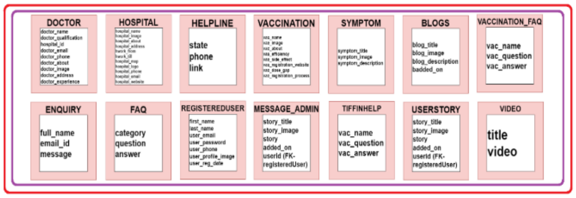

Figure 6: DJANGO Table Schema having variable names

The Figure 7 describes the architecture of the project. The project is mainly divided into two parts Django admin and User side web app. The Django admin contains the Tables and the web app contains the webpages and the templates. The app contains a combination of both static and dynamic webpages.

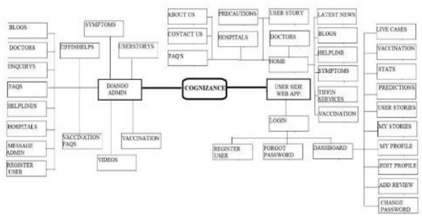

Figure 7: COGNIZANCE architecture

VI. RESULTS AND DISCUSSION

The web application contains numerous webpages. COGNIZANCE is combination of both static and dynamic webpages. Figure 8 demonstrates the interface of the application. The web app features a comprehensive footer with essential virus information and quick links. The footer illustrates contact details for the developer. The homepage header displays a counter for total deaths, confirmed cases, and recoveries, along with helpline info. It showcases three new blogs, hospitals, FAQs, user stories, doctors, and tiffin services. Washing hands guidance is at the bottom. A top counter tracks deaths, cases, and recoveries. The site delves into COVID-19, its symptoms, prevention, and nutrition. The Contact Us page houses a user form and Google map. Blogs boost site appeal with "READ MORE" links, while FAQs and helplines for Indian states and UTs are covered. Preventive measures and handwashing guidelines are detailed. Notably, common COVID-19 symptoms are presented in an attractive, user-friendly format, allowing the admin to update them as needed.

Figure 8: COGNIZANCE interface

For registered users, entering a valid registered email and password is required to log in. To register, users must provide their First Name, Last Name, Email, and password. Upon registration, a welcome email is sent to the user. In case a registered user forfeits their password, they can retrieve it by entering their registered email and clicking "Send password," which is sent to their registered email.The Dashboard as shown in Figure 9 displays details of the logged-in user. It features a user-friendly sidebar with real-time COVID-19 data for India, recently added reviews, suggestions, inquiries, and user stories with titles, images, and posting dates. Users can share their COVID-19 experiences using a text editor with formatting options like color, font size, bold, and italic. Users can view and delete their published stories. The Dashboard webpage 'my profile' shows the user's filled details, including First Name, Last Name, optional Phone Number, Email, and registration date, along with a profile picture. Clicking "EDIT PROFILE" redirects to the edit profile webpage, allowing users to update their profile image and phone number. Profile pictures must be in jpg, jpeg, or png format, and the registered email, name, and registration date are fetched from the MyProfile Page.

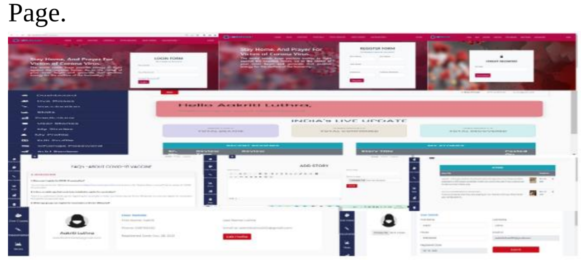

Figure 9: COGNIZANCE Dashboard interface

6.1 LIVE CASES

User has an Option to view live cases of Provinces of India as well as Worldwide Countries. In Figure 10 the user has option to select the India's cases or worldwide cases. Further user needs to select the Province/state for the India's live case and on other hand user need to select the country from the dropdown menu to check the live cases for a certain country of the world. The live data is fetched from an API.

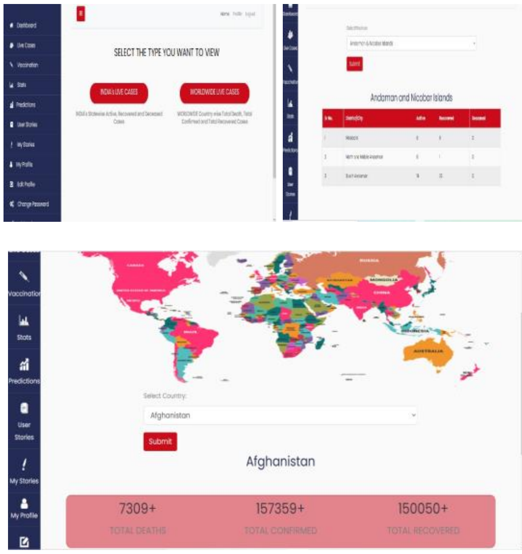

Figure 10: Live Cases Page Interface

6.2 VISUALIZATIONS

User can select the type of visualization depending on his/choice. User can Visualise data dynamically as well as statically as shown in Figure 11. In case of static visualization, the analysis is made according to the defined question and in case of dynamic the user needs to select the fields and further analysis of data are achieved based on the selected values from the dropdown menu. The visualization are made using Python's Pandas and Matplotlib library. Some examples of the analysis are shown in Figure 12, Figure 13,Figure 14 and Figure 15. And such kinds of analysis are made on the data.

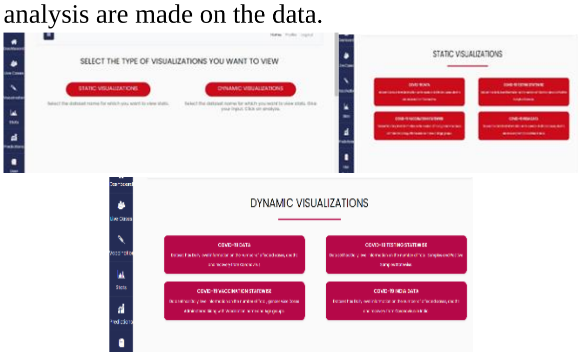

Figure 11: Visualisation Page Interface

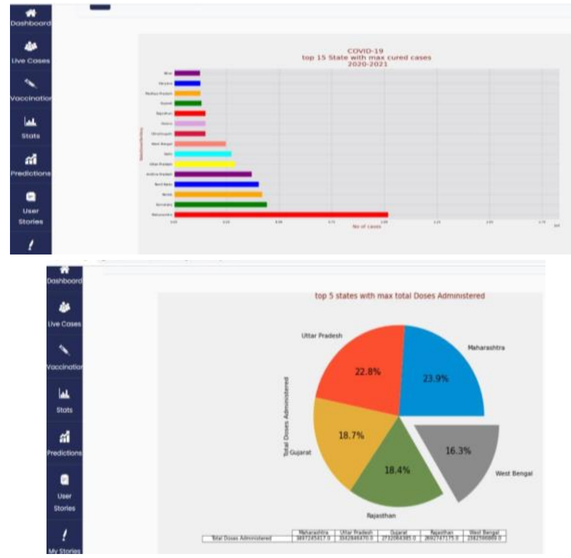

Figure 12: Visualisation displaying top 15 states with max cured cases and Top 5 states with max total doses administered

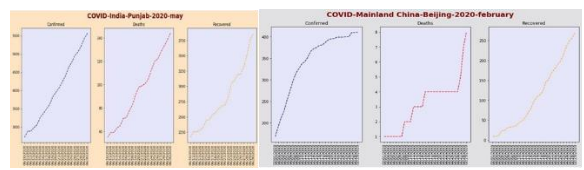

Figure 13: Visualisation displaying the Covid cases in Punjab- India in may-2020 and Beijing- China in Februrary-2020

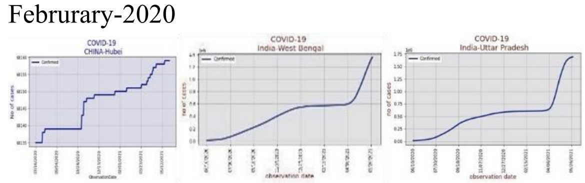

Figure 14: Visualisation displaying the Confirmed COVID-19 cases in Hubei- CHINA, West Bengal-INDIA, Uttar Pradesh- INDIA in 2020-2021

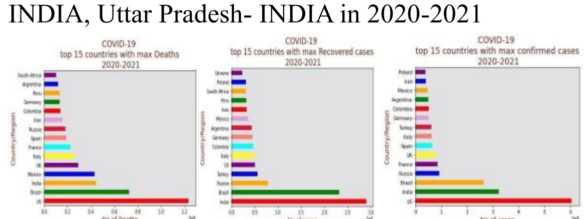

Figure 15: Visualisation displaying the top 15 Countries with maximum Deaths, Recovered cases and Confirmed cases due to COVID19 in 2020-2021

6.3 PREDICTION

The user can view the predictions related to COVID-19 testing ,cases, vaccinations and also predict whether the user is COVID positive or negative depending upon the symptoms as shown in Figure 16.

Figure 17, Figure 18, Figure 19 and Figure 20 contain a few examples of predictions performed on data.

Figure 16: Prediction Page Interface

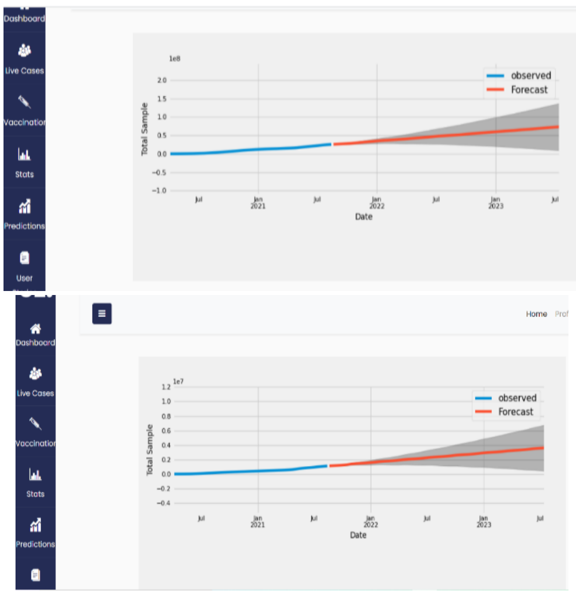

Figure 17: Prediction displaying India's statewise Testing in Andhra Pradesh and Goa

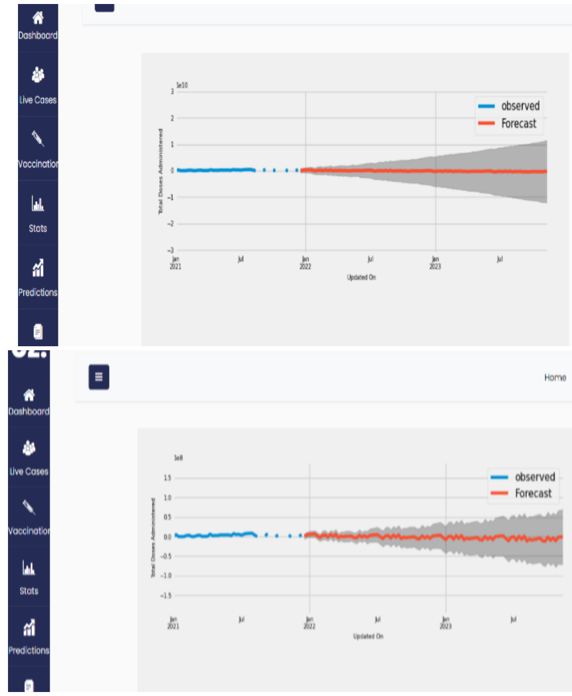

Figure 18: Prediction displaying India's Total Doses Administered and Total Doses Administered in Punjab

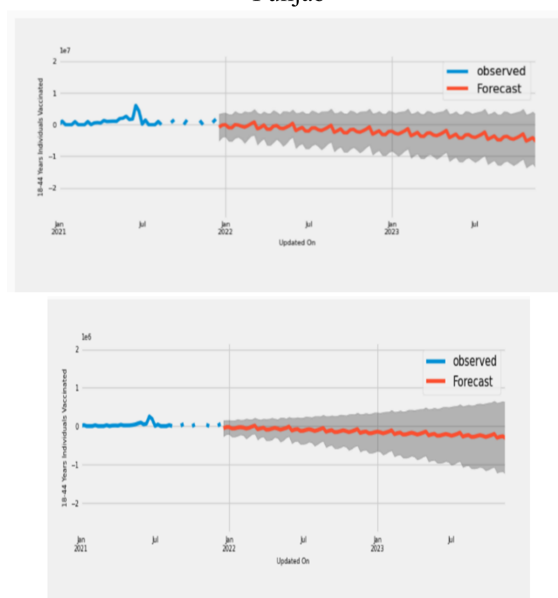

Figure 19: Prediction displaying 18-44 years individuals vaccinated in Gujrat and Goa

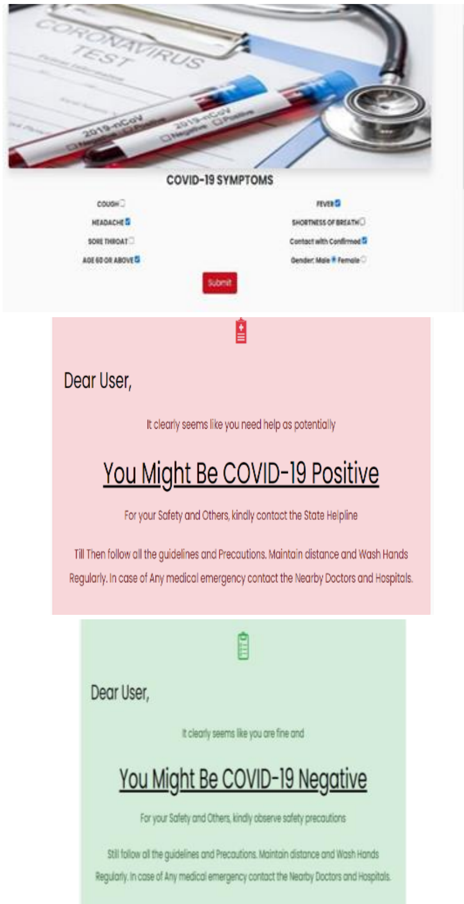

Figure 20: Prediction displaying result for covid-19 Symptoms

CONCLUSION

Since 2019, the world has been plagued by a devastating illness known as the coronavirus (Covid-19). A pandemic has been identified by WHO as a result of its death-dealing characteristics. China serves as the virus's focal point, from which it spreads slowly through human contact around the world. To restrict the virus's growth, the administrations of all nations have implemented a number of stringent measures, including full or partial lockdowns. The study suggests the future forecasting of COVID-19 by utilizing Python and machine learning. The system was developed using the Django web framework with the usage of HTML, CSS, and Bootstrap. The system would help in maintaining Covid-19 protocol and precautions and patients would be able to get proper and on-time treatment and required information thus saving many lives. In addition, now during the pandemic, the system will help users to get medical services like consultation with doctors and nearby hospitals.

REFERENCES

[1] Pradnyana G A and Darmawiguna G M 2021 Web-Based System for Bali Tourism Sentiment Analysis during The Covid-19 Pandemic using Django Web Framework and Naive Bayes Method. *Advances in Social Science, Education and Humanities Research* 613: 316-320. https://doi.org/10.2991/assehr.k.211222.050

[2] Pourhomayoun M and Shakibi M 2021 Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health 20:100178.* http://doi.org/10.1016/j.smhl.2020.100178

[3] Alves M A, Castro G Z, Oliveira B A S, Ferreira L A, Ramírez J A, Silva R, et al. 2021 Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs. *Computers in Biology and Medicine* 132: 104335. https://doi.org/10.1016/j.compbiomed.2021.104335

[4] Hussain A H, Bassam N A, Zayegh A and Ghawi S A 2022 Prediction and evaluation of healthy and unhealthy status of COVID-19 patients using wearable device prototype data. *MethodsX* 9:101618. https://doi.org/10.1016/j.mex.2022.101618

[5] Villavicencio C N, Macrohon J J, Inbaraj X A, Jeng J H and Hsieh J G 2022 Development of a Machine Learning Based Web Application for Early Diagnosis of COVID-19 Based on Symptoms. *Diagnostics* 12(4):821. https://doi.org/10.3390/diagnostics12040821

[6] Kumar G K, Raza K and Priyanka 2022 COVID-19 HOSPITAL MANAGEMENT SYSTEM USING DJANGO FRAMEWORK. *International Journal of Creative Research Thoughts (IJCRT)* 10: d463-d466. www.ijcrt.org, IJCRT2205385, ISSN: 2320-2882 https://ijcrt.org/papers/IJCRT2205385.pdf

[7] Miralles-Pechuán L, Kumar A and Suárez-Cetrulo A L 2023 Forecasting COVID-19 cases

using dynamic time warping and incremental machine learning methods. *Expert Systems 40(6):e13237*. https://doi.org/10.1111/exsy.13237

[8] Raman G, Ashraf B, Demir Y K, Kershaw C D, Cheruku S, Atis M, et al. 2023 Machine learning prediction for COVID-19 disease severity at hospital admission. *BMC Medical Informatics and Decision Making* 23(1):46. https://doi.org/10.1186/s12911-023-02132-4

[9] Paul S G, Saha A, Biswas A A, Zulfiker M S, Arefin M S, Rahman M M, et al. 2023 Combating Covid-19 using machine learning and deep learning: Applications, challenges, and future perspectives. *Array* 17:100271. https://doi.org/10.1016/j.array.2022.100271

[10] Boateng A, Maposa D, Mokobane R, Darikwa T and Gyamfi C 2023 Analysis of COVID-19 cases and comorbidities using machine learning algorithms: A case study of the Limpopo Province, South Africa. *Scientific African* 21: e01840. https://doi.org/10.1016/j.sciaf.2023.e01840

[11] https://www.kaggle.com/

[12] Cui J, Li F and Shi Z L 2019 Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology* 17(3): 181-192. https://doi.org/10.1038/s41579-018-0118-9

[13] Rasheed J, Jamil A, Hameed A A, Al-Turjman F and Rasheed A 2021 COVID-19 in the Age of Artificial Intelligence: A Comprehensive Review. *Interdisciplinary Sciences – Computational Life Sciences* 13(2): 153-175. https://doi.org/10.1007/s12539-021-00431-w

[14] Shoko C and Njuho P 2023 ARIMA MODEL IN PREDICTING OF COVID-19 EPIDEMIC FOR THE SOUTHERN AFRICA REGION. *African Journal of Infectious Diseases* 17(1):1–9. https://doi.org/10.21010/Ajidv17i1.1

[15] Sujath R, Chatterjee J M and Hassanien A E 2020 A machine learning forecasting model for COVID-19 pandemic in India. *Stochastic Environmental Research and Risk Assessment* 34(7): 959–972. https://doi.org/10.1007/s00477-020-01827-8

[16] [16] Sparks M A, South A M, Badley A D, Baker-Smith C M, Batlle D, Bozkurt B, et al. 2020 Severe Acute Respiratory Syndrome Coronavirus 2, COVID-19, and the Renin-Angiotensin System: Pressing Needs and Best Research Practices. *Hypertension* 76(5): 1350-1367. https://doi.org/10.1161/HYPERTENSIONAHA.120.15948

[17] Jones C W, Adams A C, Murphy E, King R P, Saracco B, Stesis K R, et al. 2021 Delays in reporting and publishing trial results during pandemics: cross sectional analysis of 2009 H1N1, 2014 Ebola, and 2016 Zika clinical trials. *BMC Medical Research Methodology* 21(1):120. https://doi.org/10.1186/s12874-021-01324-8

[18] Alabdulrazzaq H, Alenezi M N, Rawajfih Y, Alghannam B A, Al-Hassan A A and Al-Anzi F S 2021 On the accuracy of ARIMA based prediction of COVID-19 spread. *Results in Physics* 27:104509. https://doi.org/10.1016/j.rinp.2021.104509