**COVID 19 VACCINE ANALYSIS**

**DESCRIPTION:**

The project involves analyzing COVID-19 data to gain insights into the impact of vaccination campaigns on public health. We aim to understand vaccination rates, vaccine efficacy, and the relationship between vaccination and the spread of COVID-19.

**DATASET AND ITS DETAIL (WWW.KAGGLE.COM/DATA):**

The dataset was obtained from Kaggle, a well-known data science and machine learning platform. It provides a comprehensive COVID-19 dataset with information on cases, vaccinations, demographics, and other related variables.

LINK: https://www.kaggle.com/datasets/swatikhedekar/state-wise-india-covid19vaccination

**DETAILS ABOUT COLUMNS**

The columns are;

* State/UTs
* Total Vaccination Doses
* Dose 1
* Dose 2
* Population

**State/UTs :**

* The "State/UTs" column in the dataset refers to the Indian states and union territories (UTs) where COVID-19 vaccination data is recorded.
* Each entry in this column specifies the specific region or administrative division in India where the vaccination data is associated.

**Total Vaccination Doses :**

* The "Total Vaccination Doses Administered" column in the dataset represents the cumulative count of COVID-19 vaccine doses administered within each Indian state or union territory.
* This count includes all doses given, including both the first dose (initial vaccination) and the second dose (booster or follow-up vaccination), if applicable.

**Dose 1:**

* The "Dose 1" column in the dataset represents the number of individuals who have received the first dose of a COVID-19 vaccine within each Indian state or union territory.

**Dose 2:**

* The "Dose 2" column in the dataset represents the number of individuals who have received the second dose (booster or follow-up dose) of a COVID-19 vaccine within each Indian state or union territory.

**Population:**

* The "Population" column in the dataset provides the estimated population of each Indian state or union territory (UT).
* This figure represents the total number of residents in each region and serves as a fundamental demographic statistic.
* It's a crucial reference point for assessing COVID-19 vaccination coverage in relation to the population size of each state or UT.
* Comparing vaccination data to population data helps in understanding the proportion of people who have been vaccinated relative to the total population, which is vital for evaluating the effectiveness and reach of vaccination efforts in different regions of India.

**Tools and Libraries**

* **Python:** Utilize Python as the primary programming language.
* **Data Analysis:** Pandas for data manipulation, NumPy for numerical operations.
* **Data Visualization:** Matplotlib, Seaborn, Plotly for creating visualizations.
* **Geospatial Analysis:** Geographic Information System (GIS) tools like GeoPandas.
* **Machine Learning:** Scikit-learn for predictive modeling.
* **Web Development:** Flask or Django for building interactive dashboards.

**Library Installation:**

* To install the necessary libraries, you can use Python's package manager, pip. Open your command line or terminal and run the following commands
* pip install pandas requests matplotlib seaborn numpy geopandas folium scikit-learn plotly dash streamlit

**TRAIN AND TEST:**

1. **Data Loading:**

Start by importing the dataset from Kaggle using Python and Pandas**.**

1. **Data Preprocessing:**

Clean and prepare the dataset, addressing missing values and performing feature engineering if necessary.

1. **Data Splitting:**

Divide the data into training and testing sets, with the training set usually being larger (e.g., 80% for training and 20% for testing).

1. **Model Selection:**

Choose a machine learning model suitable for your analysis, such as regression or decision trees.

1. **Model Training:**

Train the selected model using the training data.

1. **Model Evaluation:**

Assess the model's performance on the testing data using metrics like Mean Absolute Error (MAE) or R-squared (R^2).

**METRICS:**

1. **Vaccination Coverage Rate (VCR):** This metric calculates the percentage of the population that has received the COVID-19 vaccine. A higher VCR indicates a more successful vaccination campaign.
2. **False Positive Rate (FPR):** This metric assesses the accuracy of your system in identifying vaccinated individuals. It measures the percentage of individuals incorrectly classified as unvaccinated (false negatives). A lower FPR is desired to minimize false alarms.
3. **False Negative Rate (FNR):** FNR measures the percentage of vaccinated individuals who are incorrectly classified as unvaccinated (false negatives). A lower FNR is essential to ensure that vaccinated individuals are correctly identified.

**CONCLUSION:**

INNOVATIVE PHASE:

The objective of our project is to enhance public health protection by analysing data on vaccinated individuals. In the event of a future COVID-19 outbreak, our project aims to leverage our dataset on vaccination records to accurately identify and isolate individuals who have not been vaccinated. This proactive approach is designed to help mitigate the spread of the virus and safeguard the health of the community. To assess the effectiveness of our project, we will utilize a range of metrics, including vaccination coverage rate, false positive rate, false negative rate, precision, recall, F1 score, accuracy, specificity, and relevant area-under-the-curve measurements to ensure the accurate identification of vaccinated and unvaccinated individuals.

**K.Kirubashree**

**Nandha college of technology**

**11.10.23**