NAME: CHITTOOR GANESH

REGISTER NUMBER: 713921104006

NAAN MUDHALVAN ID: au713921104006

EMAIL ID: chittoorganesh37@gmail.com

TITLE: COVID-19 VACCINES ANALYSIS

Introduction

COVID-19 vaccines have been highly effective in preventing severe illness, hospitalization, and death from the virus. However, there is still much to learn about how to optimize vaccine deployment strategies to maximize their impact. This paper will conduct an in-depth analysis of COVID-19 vaccine data, including efficacy, distribution, and adverse effects, to provide insights that can aid policymakers and health organizations in making informed decisions about vaccine deployment.

The development and distribution of COVID-19 vaccines have been pivotal in the global fight against the pandemic caused by the novel coronavirus, SARS-CoV-2. As vaccines became widely available, their effectiveness, safety, distribution, and impact on public health became subjects of intense interest and scrutiny. COVID-19 vaccine analysis encompasses a broad range of research and data-driven investigations aimed at better understanding these vaccines and optimizing their deployment.

Dataset are gathered from following link:

Link: https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress

Data is collected daily from **Our World in Data** GitHub repository for covid-19, merged and uploaded. Country level vaccination data is gathered and assembled in one single file. Then, this data file is merged with locations data file to include vaccination sources information. A second file, with manufacturers information, is included.

Detail About Columns:

The data (country vaccinations) contains the following information:

- **Country**- this is the country for which the vaccination information is provided;
- **Date** - date for the data entry; for some of the dates we have only the daily vaccinations, for others, only the (cumulative) total;
- **Total number of vaccinations** - this is the absolute number of total immunizations in the country;
- **Total number of people vaccinated** - a person, depending on the immunization scheme, will receive one or more (typically 2) vaccines; at a certain moment, the number of vaccination might be larger than the number of people;
- **Daily vaccinations** - for a certain data entry, the number of vaccination for that
- **Vaccines used in the country** - total number of vaccines used in the country (up to date);

Numpy:

NumPy is the fundamental package for scientific computing in Python.it is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation

Pandas

Pandas is a Python library used for working with data sets.It has functions for analyzing, cleaning, exploring, and manipulating data. Pandas allows us to analyze big data and make conclusions based on statistical theories. Pandas can clean messy data sets, and make them readable and relevan

Matplotlib

Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. It was introduced by John Hunter in the year 2002

Seaborn

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

OS(operating system)

Python has a built-in os module with methods for interacting with the operating system, like creating files and directories, management of files and directories, input, output, environment variables, process management.

way to download in python libraries

Any Python library can be installed manually in just one step using the command: python3 <FILE_NAME>. py install. The command pip install package_name is used in Python to install libraries using scripts.

HOW TO TRIAN AND TEST ?

- Data Collection and Preprocessing:
- Data Exploration:
- Feature Engineering,
- Data Splitting:
- Model Selection:
- Model Training,
- Hyperparameter Tuning:,
- Model Evaluation:
- Interpret Results:
- Visualization and Reporting:
- Regular Updates:
- Ethical Considerations:

REST OF EXPLANATION:

Hypothesis Testing:

If your analysis aims to answer specific research questions or hypotheses, consider performing hypothesis testing. This involves statistical tests to determine whether observed differences or relationships in the data are statistically significant.

Time Series Analysis:

COVID-19 vaccination data often includes a temporal dimension. Use time series analysis techniques to understand vaccination trends over time, detect seasonality, and assess the impact of interventions or policies.

Geospatial Analysis:

If your dataset includes geographic information, perform geospatial analysis to identify regional vaccination disparities and their effects on infection rates. Tools like Geographic Information Systems (GIS) can be valuable for this purpose.

Cohort Analysis:

Divide the dataset into different cohorts based on demographic variables (e.g., age, gender), comorbidities, or other relevant factors. Analyze how vaccination impacts different subgroups within the population.

Predictive Modeling:

If your goal is to predict future vaccination trends or infection rates, consider using predictive modeling techniques. Time series forecasting, regression, and machine learning models can be used for this purpose.

ACCURACY CHECK:

Classification Metrics (e.g., Vaccine Efficacy Analysis):

Accuracy: Measures the proportion of correct predictions in a binary classification scenario (vaccinated vs. unvaccinated).

Precision: Measures the proportion of true positive predictions (correctly identified vaccinated individuals) out of all positive predictions.

Recall (Sensitivity): Measures the proportion of true positive predictions out of all actual vaccinated individuals.

Specificity: Measures the proportion of true negative predictions (correctly identified unvaccinated individuals) out of all actual unvaccinated individuals.

F1-Score: A harmonic mean of precision and recall, which balances both metrics. It's particularly useful when there is an imbalance between vaccinated and unvaccinated groups.

Regression Metrics (e.g., Predicting Infection Rates):

Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual values. Lower MAE indicates better accuracy.

Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values. MSE is useful for giving more weight to larger errors. Root Mean Squared Error (RMSE): The square root of MSE, it's in the same unit as the target variable and provides better interpretability.