**NAAN MUDHALVAN-IBM DATA ANALYTICS WITH COGNOS**

**PROJECT PHASE 2 INNOVATION**

**PROJECT TITLE:**

***COMPREHENSIVE ANALYSIS OF COVID-19 VACCINATION DATA:***

Enhancing deployment strategies for optimal public health impact

**Team Members**

|  |  |  |
| --- | --- | --- |
| **NAMES** | **ROLL.NO** | **E-Mail id** |
| Shanjanaa.G | 2021115099 | [*shanju.jaya@gmail.com*](mailto:shanju.jaya@gmail.com) |
| Shantha priya.M | 2021115100 | [*mshanthapriya@gmail.com*](mailto:mshanthapriya@gmail.com) |
| Sharmitha.J | 2021115101 | [*sharmithaj23@gmail.com*](mailto:sharmithaj23@gmail.com) |
| Ramyaa.R | 2021115318 | [*ramyarajaraman31@gmail.com*](mailto:ramyarajaraman31@gmail.com) |
| Saravanakumar.P | 2021115319 | [*sksaravana012@gmail.com*](mailto:sksaravana012@gmail.com) |

***INTRODUCTION***

In this phase, we build upon our established objectives and delve into advanced techniques for COVID VACCINE ANALYSIS. Our focus lies in harnessing cutting-edge machine learning algorithms to predict vaccination trends and extract valuable insights from public health data and user feedback. By optimizing vaccine distribution and resource allocation, our goal is to enhance public health outcomes and ensure a safer future for all.

***DATA INTEGRATION***

Essential in COVID VACCINE ANALYSIS, data integration involves collecting, transforming, and consolidating data from varied sources like health databases and vaccination centers. This process creates a unified dataset, ensuring accurate insights and informed decision-making for optimizing vaccine distribution and public health strategies.

***DATA PREPARATION***

In COVID VACCINE ANALYSIS, data preparation is crucial. Raw data from diverse sources, such as health databases and vaccination centers, undergoes meticulous cleaning, transformation, and structuring. The focus is on rectifying inconsistencies, handling missing values, and standardizing formats. Filtering out irrelevant or erroneous information results in a clean dataset, ensuring the reliability and accuracy of insights for informed decision-making in public health strategies**.**

***DATA MODELING***

In COVID VACCINE ANALYSIS, data modeling is a pivotal step. Prepared data is structured within IBM Cognos Framework Manager, defining relationships, key metrics, and enabling efficient querying. Specifically, a time-based dimension is essential for trend analysis. This structured model forms the foundation for generating insightful reports and dashboards, simplifying data retrieval and ensuring users can access relevant metrics and KPIs efficiently for informed decision-making in public health strategies.

***DESIGNING REPORTS:***

In COVID VACCINE ANALYSIS, report design is vital. Utilizing IBM Cognos tools, data insights are transformed into visually appealing presentations, employing appropriate visualizations for KPIs and metrics. Effective design ensures clarity and relevance, enhancing user understanding. Tailored reports provide a dynamic platform for data-driven decision-making in public health strategies.

***ANALYSIS***

In the COVID VACCINE ANALYSIS phase, data is examined to unveil patterns and user behavior, guiding actions like content optimization and marketing strategy adjustments. These insights drive iterative improvements, ensuring alignment with organizational goals and enhancing user satisfaction in public health strategies.

**CLUSTURING ALGORITHM IN MACHINE LEARNING:**

Clustering is a type of unsupervised machine learning technique that involves grouping similar data points together in order to discover underlying patterns or structures within a dataset. Clustering algorithms aim to partition the data into clusters or groups where data points within the same cluster are more similar to each other than to those in other clusters.

**K-Means Clustering Algorithm**

K-Means clustering is a popular unsupervised machine learning algorithm that divides data into K clusters based on the nearest mean. It initializes centroids, assigns data points iteratively, and updates centroids until convergence. This method, efficient for tasks like customer segmentation and document categorization, requires specifying the number of clusters beforehand.

**Hierarchical Clustering Algorithm**

Hierarchical Clustering is an unsupervised ML technique creating a cluster hierarchy. It can be agglomerative (bottom-up) or divisive (top-down), merging or dividing clusters until a dendrogram reveals hierarchical data relationships, aiding insight into cluster structures.

**DBSCAN Clustering Algorithm**

DBSCAN is a density-based clustering algorithm in machine learning that groups data points based on their density, eliminating the need for specifying cluster numbers beforehand. It identifies clusters as high-density areas separated by lower-density regions, and it can label outliers as noise. Robust to irregular shapes and variable densities, DBSCAN is ideal for datasets with complex cluster shapes and varying data point densities.

**Gaussian Mixture Model(GMM)**

Gaussian Mixture Model (GMM) is a probabilistic model used in machine learning for clustering and density estimation. It identifies multiple Gaussian distributions within data, assigns data points probabilistically, and is adept at modeling complex data distributions, commonly applied in tasks like image segmentation and speech recognition.

**Fuzzy Clustering**

Fuzzy clustering, like Fuzzy C-Means (FCM), assigns data points degrees of membership to multiple clusters, allowing for simultaneous association with multiple clusters. It's valuable for data with unclear boundaries or uncertainty about cluster assignments, commonly used in tasks like image and customer segmentation with overlapping preferences.

***Time Series Forecasting Models for COVID Vaccine Analysis:***

**ARIMA (AutoRegressive Integrated Moving Average)**

ARIMA models are adept at predicting time series data, making them suitable for forecasting trends in COVID vaccine-related metrics, such as vaccination rates or case counts.

**Exponential Smoothing**

Exponential smoothing methods like Holt-Winters are effective for capturing seasonality and trends in time series data, making them valuable for analyzing and predicting patterns in vaccine distribution and public health metrics.

***CONCLUSION***

Here, we have discussed the project design that we will follow and the machine learning models we will utilize in this project