**Project**: Application of DBSCAN in the international Humanitarian NGO

By: Suriya Narayanan

**Introduction:**

In this project, we aim to apply the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to enhance the data analysis capabilities of an International Humanitarian NGO. By utilizing DBSCAN, we seek to identify meaningful clusters and outliers in humanitarian data, enabling the NGO to optimize resource allocation and respond more efficiently to crises. The project's outcomes can help International Humanitarian NGO to extend their help to save lives, reduce suffering, and improve the overall effectiveness of humanitarian interventions for countries, making data-driven decisions an integral part of their operations.

**Problem statement:**

After the current funding programs, International Humanitarian NGO has raised around $ 10 million. Now the CEO of the NGO needs to decide how to use this money strategically and effectively. The significant issues that come while making this decision have mostly related to the countries in dire need of Aid. Our job is as Data analyst is to classify the countries using the socio-economic and health factors that determine the overall development of nations. After this analysis, we need to suggest countries that the CEO needs to focus on and give the highest priority.

**Process:**

Data Collection: Collect relevant datasets from various sources, including satellite imagery, surveys, social media, and previous mission reports. Ensure data quality and consistency for accurate analysis.

Data Preprocessing: Clean and preprocess the data to handle missing values, outliers, and inconsistencies. Transform the data into a suitable format for DBSCAN algorithm implementation.

DBSCAN Algorithm Implementation: Implement the DBSCAN algorithm to identify clusters and outliers in the humanitarian data. Tune the algorithm's parameters (epsilon and minPts) to fit the characteristics of the dataset.

Spatial Analysis: Visualize the clustered data on maps to identify areas with high concentrations of vulnerable populations, infrastructural damage, or resource scarcity. This analysis will aid in resource allocation and response planning.

**Output:**

**label -1 specifies medium child mortality,medium exports,high income,medium inflation and medium gdp**

**label 0 specifies high child mortality,low exports,low income,high inflation and low gdp**

**label 1 specifies low child mortality,high exports,medium income,low inflation and medium gdp**

**Reference:**

1. [How Does DBSCAN Clustering Work? | DBSCAN Clustering for ML (analyticsvidhya.com)](https://www.analyticsvidhya.com/blog/2020/09/how-dbscan-clustering-works/#:~:text=DBSCAN%20is%20a%20density%2Dbased,points%20into%20a%20single%20cluster.)