**‘Naadhagama’ Geospatial Analysis System Implementation Guide**

**1. Introduction.**

**1.1 Purpose of the System.**

The Geospatial Analysis System for the Naadhagama Music Concert is designed to provide insights into the geographic distribution of Instagram posts related to the event. It incorporates data collection, preprocessing, analysis, and predictive modeling to help event organizers monitor user engagement, identify patterns, and predict future social media activity.

**2. Getting Started**

**2.1 Prerequisites**

Ensure you have the required Python libraries installed.

* Pandas
* Numpy
* Geopandas
* Matplotlib
* Seaborn
* Geopy
* Dash
* Plotly
* Folium
* scikit-learn
* statsmodels

Ensure extracted dataset is available. Data is extracted by Apify Instagram Hashtag Scraper API.

**2.2 Installation**

1. Clone the repository containing the implementation files.
2. Install required Python libraries mentioned above.
3. Open **‘task\_4\_geospatial\_application.ipynb’** in a Jupyter Notebook to execute the implementation.

**3. Data Collection and Preprocessing**

**3.1 Data Source**

Data for the Naadhagama Music Concert is collected from Instagram using the Apify Instagram Hashtag Scraper API. Key columns are extracted, and preprocessing involves cleaning, deduplication, and imputation.

**4. Data Analysis**

**4.1 Spatial Distribution Analysis**

Interactive maps using Folium display the spatial distribution of Instagram posts related to Naadhagama, providing a visual overview.

**4.2 Hotspot Analysis**

Identify areas with high user engagement by analyzing likes and comments, helping organizers focus on content that resonates with the audience.

**4.3 Spatial Cluster Analysis**

K-Means clustering identifies district-wise patterns, aiding in understanding how user engagement varies across different regions.

**4.4 Temporal Analysis**

Examine posting time patterns to identify peak activity periods, facilitating strategic planning for future events.

**5. Predictive Modeling**

**5.1 Model Development**

The historical monthly posts counts are considered for forecasting people engagement for Naadhagama events in Year 2024. ARIMA model is used for time series forecasting.

**5.2 Model Evaluation**

Model performance is evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

**5.3 Email Notification**

An email notification mechanism is implemented to alert users about predicted future engagement.

**6. Dashboard Overview**

**6.1 Dashboard Components**

The dashboard integrates interactive maps, key metrics, and predictive model outputs to provide a holistic view of Naadhagama concert.

**6.2 Accessing the Dashboard**

Access the dashboard by opening ‘**task\_4\_geospatial\_application.ipynb’** in a Jupyter Notebook and executing the provided code cells.

To open the dashboard, hit the localhost URL in your browser - http://127.0.0.1:8050/