
RESEARCH ATTACHMENT PROGRAM

Disease mapping using AI cluster

By : Sammed M

TOPICS TO BE DISCUSSED

- 1) Introduction
- 2) AI Cluster
- 3) DBSCAN
- 4) Neural Network
- 5) Cluster Analysis
- 6) Matplotlib
- 7) Advantages AI cluster in disease mapping.
- 8) Sample Code
- 9) Output
- 10) Conclusion

INTRODUCTION

What is AI cluster?

An AI cluster is a network of interconnected computers optimized for high-performance AI tasks. Utilizing specialized hardware like GPUs or TPUs, these clusters accelerate computations for training and inference. Managed by cluster management software, they enable efficient collaboration and scalability in tackling complex AI problems.

Why AI Clustering for Disease Mapping:

- Identifies patterns and trends in disease spread.
- Handles large and complex datasets.
- Provides precise and actionable insights for early detection and intervention.

AI Cluster

AI clustering is a machine learning technique that groups similar data points based on their inherent characteristics. It leverages data to discover patterns and structures, enabling machines to analyze and categorize data without supervision. AI clustering has various applications, including customer service automation, product development, and supply chain management, where it helps uncover hidden patterns and trends in vast datasets.

Modeling: involves selecting an appropriate clustering algorithm, such as K-Means, Hierarchical Clustering, or DBSCAN, based on the specific problem and data characteristics. The algorithm is then trained on the data, and the resulting clusters are evaluated for their quality, interpretability, and relevance to the problem at hand.

Applications:

- City Planning Optimization Strategies
- Customized Training Sets Created
- Advanced Threat Detection Systems
- Personalized Advertising Campaigns Developed

DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a popular clustering algorithm used in data mining and machine learning. It groups together data points that are closely packed, distinguishing dense regions from sparse ones, without requiring the number of clusters to be specified beforehand.

Features: DBSCAN offers flexibility in identifying clusters of varying shapes and sizes, making it robust against noise and outliers. It utilizes parameters such as epsilon (ϵ) and minimum points (MinPts) to define clusters based on density.

Implementation: DBSCAN works by iteratively exploring data points, classifying them as core points, border points, or noise points. It forms clusters by connecting core points and their directly reachable neighboring points, while noise points remain unassigned.

Advantages: DBSCAN is effective in discovering clusters of arbitrary shapes and handling data with varying densities. It automatically identifies outliers as noise points, providing insights into data quality. Moreover, it does not require the pre-specification of the number of clusters, making it suitable for exploratory data analysis.

Applications: DBSCAN finds applications in various fields such as spatial data analysis, image processing, anomaly detection, and customer segmentation. It is particularly useful in disease mapping, where it can identify spatial clusters of

NEURAL NETWORK

Image Recognition: Neural networks, especially Convolutional Neural Networks (CNNs), have revolutionized the field of computer vision by enabling machines to recognize and analyze images.

Object Detection: Neural networks are capable of identifying and detecting objects within an image, which is a crucial application of image processing.

Image Segmentation: Neural networks can convert an image into a collection of regions of pixels represented by a mask or a labelled image, a process known as image segmentation.

Image Generation: Neural networks can generate new images based on certain criteria or characteristics, learning from a large dataset of images.

Hardware Implementation: The implementation of these image processing techniques using hardware description languages like Verilog can lead to the development of efficient and high-performance hardware solutions

Matplotlib

Versatility: Matplotlib versatility shines through its ability to generate a wide array of plots, from basic line graphs to intricate heatmaps and 3D plots, ensuring suitability for diverse data visualization tasks.

Community Support: With a vast user base and extensive online documentation, Matplotlib offers robust community support, facilitating quick troubleshooting and learning for users at all skill levels.

Integration: Matplotlib seamlessly integrates with other Python libraries like NumPy and Pandas, allowing users to leverage its powerful plotting capabilities in conjunction with data manipulation and analysis tools for comprehensive data exploration.

Customization: Offering fine-grained control over plot aesthetics, Matplotlib empowers users to customize every aspect of their plots, including axes, legends, and annotations, ensuring that visualizations effectively convey insights tailored to specific analytical needs.

ADVANTAGES OF CLUSTER ANALYSIS IN DISEASE MAPPING

Identification of High-Risk Areas: Cluster analysis helps pinpoint geographical regions with elevated disease prevalence, aiding targeted intervention strategies and resource allocation for public health initiatives.

Understanding Disease Patterns: By analyzing spatial clusters, researchers gain insights into the spatial distribution and spread of diseases, facilitating better understanding of epidemiological trends and disease transmission dynamics.

Early Detection and Prevention: Detection of disease clusters enables early intervention measures, such as implementing surveillance systems or preventive measures in at-risk areas, leading to timely disease control and prevention efforts.

Resource Optimization: Disease mapping via cluster analysis optimizes resource allocation by directing healthcare resources, personnel, and funding to areas with the highest disease burden, ensuring efficient use of resources for effective disease management.

Support for Policy Making: Identifying disease clusters provides evidence-based support for policymaking and public health decision-making, enabling policymakers to implement targeted interventions and allocate resources effectively to mitigate disease spread and impact.

SAMPLE CODES

```
import pandas as pd
import matplotlib.pyplot as plt
import plotly.express as px
from sklearn.cluster import DBSCAN
from datetime import datetime

# Load the dataset
data = pd.read_csv('code1_dataset.csv')

# Convert the Date column to datetime
data['Date'] = pd.to_datetime(data['Date'])

# Extract year and month for temporal analysis
data['Year'] = data['Date'].dt.year
data['Month'] = data['Date'].dt.month

# Prepare data for clustering: combining spatial and temporal data
data['YearMonth'] = data['Year'].astype(str) + '-' + data['Month'].astype(str)
data_for_clustering = data[['Latitude', 'Longitude', 'Year', 'Month']]

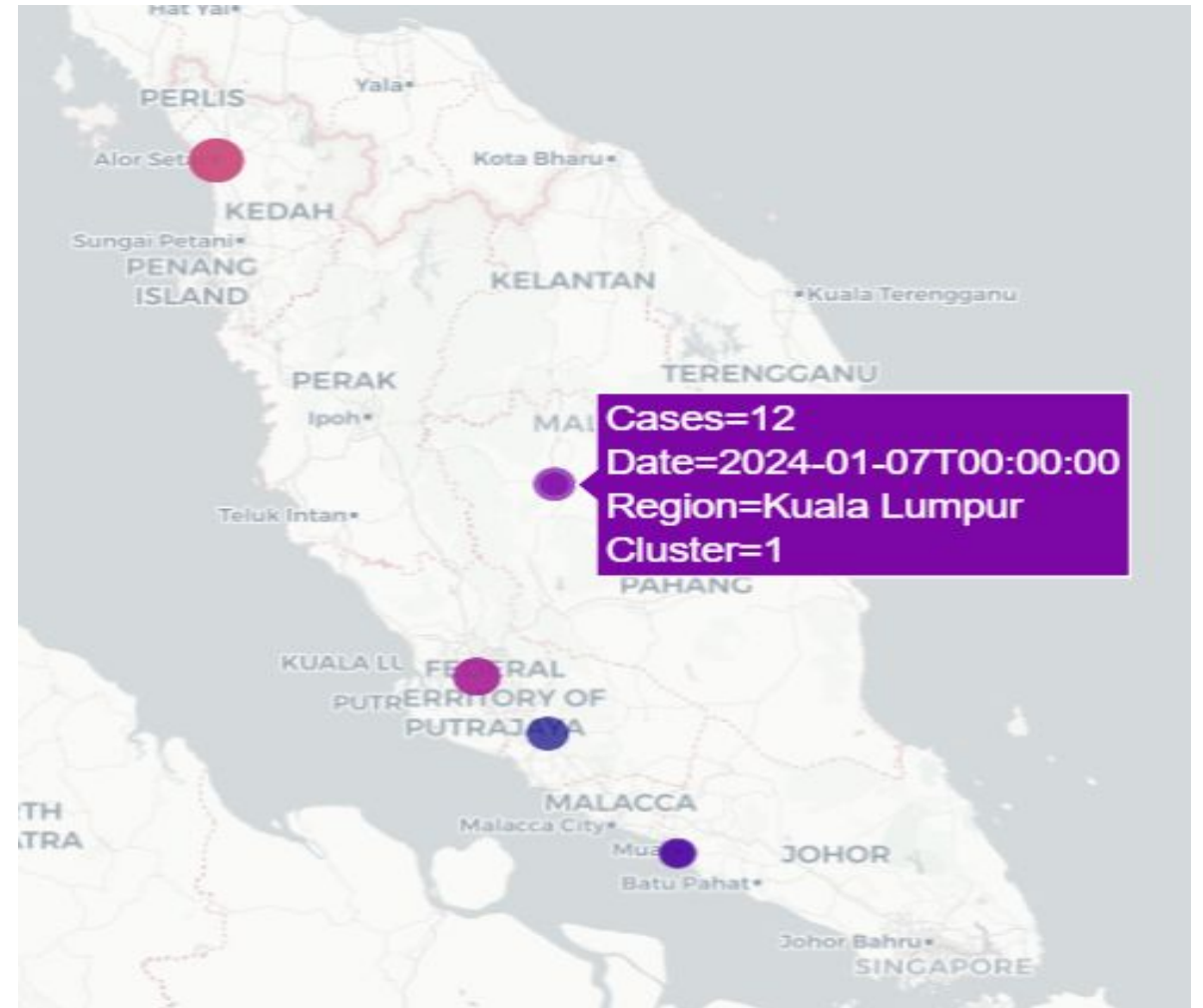
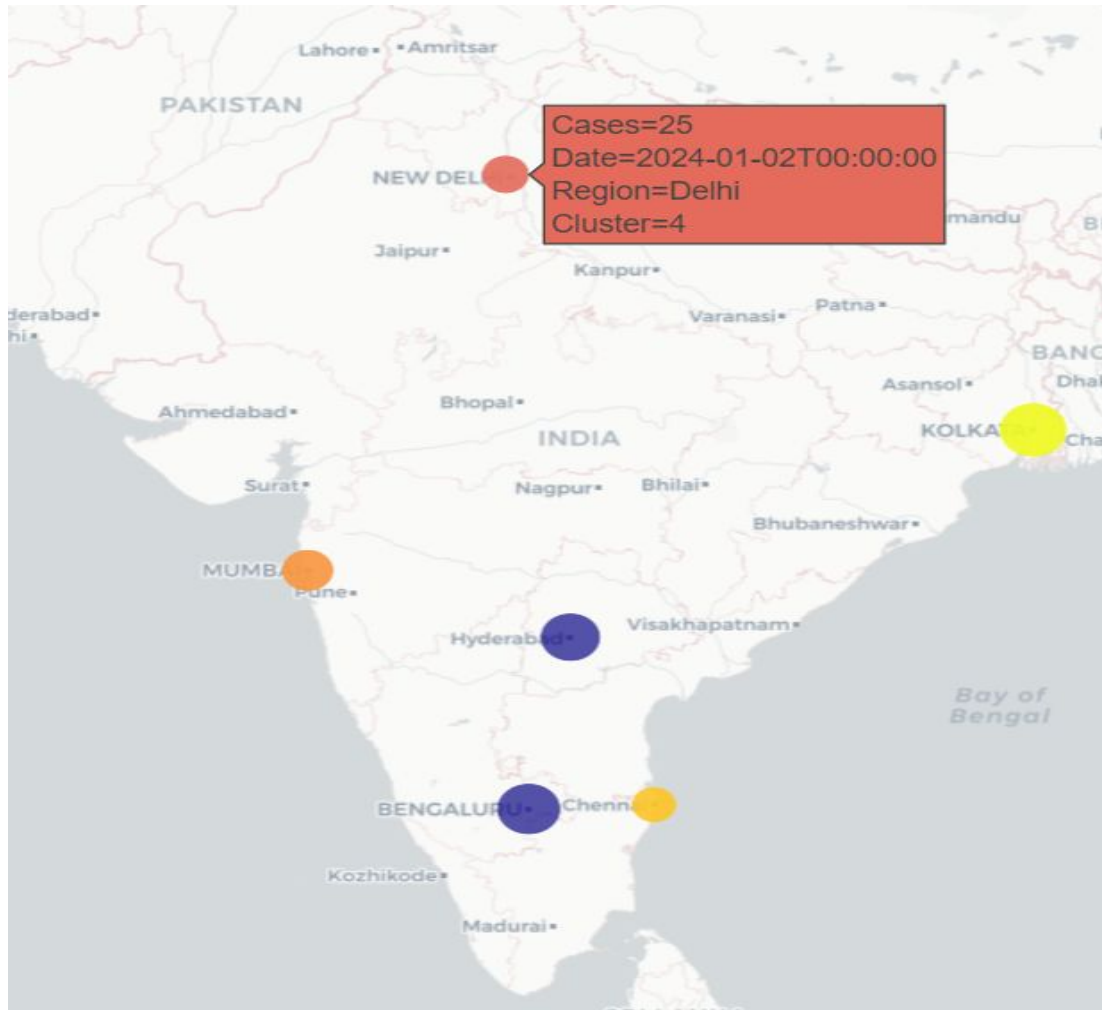
# Apply DBSCAN clustering
dbscan = DBSCAN(eps=0.3, min_samples=2)
clusters = dbscan.fit_predict(data_for_clustering)
```

```
# Add cluster labels to the dataset
data['Cluster'] = clusters

# Visualize the clusters
fig = px.scatter_mapbox(
    data, lat='Latitude', lon='Longitude', color='Cluster', size='Cases',
    hover_data={'Latitude': False, 'Longitude': False, 'Cases': True, 'Date': True, 'Region': True},
    title="Dengue Cases Clustering",
    mapbox_style="carto-positron"
)

fig.update_layout(mapbox=dict(zoom=5, center=dict(lat=3.0, lon=101.5)))
fig.show()
```

OUTPUT



CONCLUSION

In conclusion, our project on Disease Mapping using AI Clustering represents a significant advancement in public health surveillance methodologies. By utilizing AI clustering techniques, we have demonstrated the capability to analyze spatial and temporal disease patterns with accuracy and granularity. The insights from our analysis hold immense potential for informing targeted intervention strategies, resource allocation, and policy decisions aimed at lessening the impact of infectious diseases. Moving forward, our work serves as a foundation for further innovation in disease surveillance and public health, with the goal of improving health outcomes and enhancing the well-being of communities worldwide.



Thank You