Capstone Project Idea Proposal Submission

Group – 7:

1. Olana Kenea Lemesa
2. Selemon Lera Gamo
3. Yonas Nigussu
4. HaileMaryam Kebede Tsigehana
5. Heaven Alemu

Project Title

Predictive Modeling for Early Detection of Disease Outbreaks Using Epidemiological and Climate Data

Problem of Statement

Early detection of infectious disease outbreaks is crucial for public health planning and response.

Traditional surveillance methods rely heavily on manual data collection, often leading to delays in identifying potential outbreaks and taking necessary action. These delays can result in increased transmission, overwhelming healthcare systems, and higher mortality rates.
This project aims to develop an AI-driven early warning system by leveraging machine learning models trained on epidemiological and climate data. By analyzing historical patterns, environmental conditions, and population movement, the system will predict potential outbreaks before they escalate, enabling timely intervention and resource allocation.

Relevance to Sustainable Development Goals (SDGs)

This project aligns with **SDG 3:** Good Health and Well-being, which emphasizes the prevention and control of epidemics. By providing an AI-driven early warning system, this model can help minimize disease spread, improve healthcare preparedness, and ultimately save lives.

Literature Reviews

1. *A machine learning-based universal outbreak risk prediction tool - Zhang, T., Rabhi, F.,*

Chen, X., Paik, H., & MacIntyre, C. R. (2023). https://doi.org/10.1016/j.compbiomed.2023.107876 – This study explores different machine learning models and used outbreak data from 43 diseases in 206 countries to develop a universal risk prediction system that can be used across countries and diseases.

2. *Using machine learning to predict disease outbreaks and enhance public health*

surveillance - Ekundayo, N. F. (2024) https://doi.org/10.30574/wjarr.2024.24.3.3732 – This paper highlights the critical role of ML in transforming public health surveillance, focusing on its application in disease outbreak

prediction.

Our Data

- **Epidemiological Data:** Historical disease outbreak records, case reports, mortality rates (Sources: WHO, CDC, national health ministries).
- **Climate Data:** Temperature, humidity, precipitation, air quality (Sources: NASA, NOAA, local meteorological agencies).
- **Population Mobility Data (if available):** Human movement patterns derived from mobile phone usage or transport networks.
- **Data Format:** CSV, JSON, or API-based real-time data streams.
- **Preprocessing:** Data cleaning, handling missing values, feature engineering, and normalization.

Approach (Machine Learning or Deep Learning)

We will use Machine Learning (ML) models as the primary approach, with the potential to explore Deep Learning (DL) models if needed.

Since our dataset is primarily structured time-series data, we will start with ML models (Random Forest, XGBoost). If results need improvement, we will explore LSTMs (DL) for advanced time-series prediction.