



PREDICTING MALARIA OUTBREAKS IN RURAL LIBERIA USING MACHINE LEARNING

FTL Liberia Group Six (6)

ABSTRACT

This project proposes to build a machine learning model to predict malaria outbreaks in rural Liberia using environmental, demographic, and health data. The goal is to create an early warning system that helps manage resources and save lives, supporting Sustainable Development Goals (SDGs) 3 and 1.

GROUP MEMBERS

1. Nehemiah Kemayah
2. Thomas Adonis Marwolo
3. Robert Bright Yekeh
4. Joel J. Barclay

Project Background

Malaria continues to be a major health problem in rural parts of Liberia. The specific goal of this project is to build a predictive model that can forecast potential malaria outbreaks based on environmental, demographic, and health data. Early warning systems can help allocate medical resources more effectively and save lives.

Relevance to Sustainable Development Goals (SDGs)

This project supports **SDG 3: Good Health and Well-being** by aiming to reduce the impact of malaria through early detection and intervention. It also aligns with **SDG 1: No Poverty**, since malaria contributes significantly to economic hardship in affected communities, and controlling it improves economic productivity.

Literature Review

1. **Study 1:** "Predicting Malaria Incidence Using Climate Variables in Sub-Saharan Africa" – This research demonstrated the effectiveness of machine learning in forecasting malaria using rainfall, temperature, and humidity data.
2. **Study 2:** "A Machine Learning Approach for Malaria Detection and Forecasting" – This paper used health records and environmental factors to successfully predict outbreak patterns, highlighting the feasibility of using AI for disease surveillance in low-resource settings.

Project Data Description

The project will use:

- ✓ Health Data from Liberia's Ministry of Health (CSV format, weekly reports of malaria cases by region).
- ✓ Climate Data from the World Bank Climate Change Knowledge Portal (CSV format, including rainfall, temperature).
- ✓ Population Data from Liberia Institute of Statistics (Excel or CSV).

Data preprocessing steps include cleaning missing values, standardizing scales, and merging datasets by time and location.

Approach

A machine learning approach will be used due to the structured nature of the data (numerical and categorical variables) and the moderate size of the dataset. Algorithms such as Random Forests or Gradient Boosting Machines are suitable for modeling temporal-spatial patterns in malaria outbreaks, with good interpretability and relatively low computational cost compared to deep learning.