<u>MALARIA OUTBREAK PREDICTION REPORT</u>

Malaria remains one of the world's most pressing public health challenges, disproportionately affecting vulnerable populations in sub-Saharan Africa and parts of Southeast Asia. Each year, over 200 million cases are reported, leading to more than 400,000 deaths—most of them children under five. This effort aligns with **UN Sustainable Development Goal 3: Good Health and Well-being**, specifically Target 3.3, which calls for ending the epidemics of neglected tropical diseases, including malaria, by 2030. By improving predictive capabilities around malaria outbreaks, health authorities can more efficiently allocate resources—such as insecticide-treated bed nets, antimalarial medications, and rapid diagnostic tests—to communities at greatest risk.

A supervised learning approach is utilized to predict malaria cases, leveraging features including temperature, rainfall, population density, and healthcare access. The model processes data from Kaggle's "Malaria in Africa" dataset, World Bank health expenditure records, and planned NOAA weather data integration. Python libraries such as scikitlearn facilitate model development, with preprocessing steps handling missing values and data reshaping for compatibility. While the specific algorithm (e.g., Random Forest or Gradient Boosting) is under development, the supervised framework maps input features to malaria case outcomes for accurate predictions.

The system successfully integrated Kaggle's "Malaria in Africa" dataset with World Bank health expenditure data, confirming minimal missing values and robust features like temperature and rainfall for malaria prediction. Preliminary analysis suggests strong feature correlations, laying the groundwork for accurate outbreak forecasting. Once implemented, the supervised learning model is expected to reduce outbreak response times by 20–30% in high-risk African regions, enhancing resource allocation for malaria prevention.

The project prioritizes equitable health outcomes, focusing on underserved African communities disproportionately affected by malaria. However, potential biases in the dataset, such as underrepresentation of rural areas, may skew predictions, necessitating careful validation. As the data is aggregated, privacy risks are low, but accuracy across diverse sources is critical to avoid misleading health policies. Transparency is maintained through open-source code, and deployment will require collaboration with local health authorities to ensure culturally sensitive interventions and prevent over-reliance on AI without human oversight.