Forecasting Malaria Incidence Using Machine Learning

Final Year Project

Department of Computer Science

BSc. Computer Science

Author: [Your Name]

Supervisor: [Supervisor's Name]

Date: June 2025

# Abstract

Malaria remains a major public health concern in sub-Saharan Africa. This project presents a predictive model to forecast malaria case trends using historical data collected from 2019 to 2024. Machine learning techniques, specifically XGBoost, are employed to train the model. The application leverages Python for preprocessing and modeling, while Power BI serves as the dashboard platform for visualizing the forecasts. The aim is to equip health officials with a reliable tool for proactive planning and intervention.

# 1. Introduction

Malaria is a life-threatening disease affecting millions globally, especially in sub-Saharan Africa. Forecasting malaria cases can enable targeted intervention and resource allocation. This project uses data science techniques to predict future malaria trends using five years of historical data.

# 2. Problem Statement

Traditional methods of malaria management are often reactive rather than proactive. Without predictive insights, health facilities may either over-prepare or under-prepare for outbreaks. This project proposes a predictive model to estimate monthly malaria cases, enabling better preparedness.

# 3. Objectives

- Build a predictive model for monthly malaria case forecasts.

- Use real-world data to train and evaluate the model.

- Develop a dashboard using Power BI to visualize model output.

# 4. Methodology

The methodology involves several steps: data preprocessing, feature engineering, model training using XGBoost, and visualization of results. The data is cleaned and aggregated monthly. Feature engineering includes creating lag features and extracting date parts. The dataset is split into training and testing sets without shuffling to preserve the time series structure.

# 5. Tools and Technologies

- Python (pandas, xgboost, matplotlib)

- Power BI

- Microsoft Excel

# 6. Results and Evaluation

The model performance is evaluated using Mean Absolute Error (MAE). The visualized results show the predicted vs actual cases over time, demonstrating the model's ability to capture the trend with reasonable accuracy.

# 7. Limitations

The current model uses only historical case data. Environmental and demographic factors such as rainfall, temperature, and population mobility are not included. These factors could enhance prediction accuracy.

# 8. Conclusion

This project demonstrates the feasibility of applying machine learning to forecast malaria trends. The results provide useful insights for public health planning and resource allocation. Future versions of the model could integrate additional variables and be deployed as a real-time dashboard.