**AKWO MAKEMBE KING**

**CT22A016**

**Data Mining Project Overview**

**Project Title:**  
**Predictive Analysis of Malaria Outbreaks in Buea Cameroon Using Data Mining Techniques**

### 1. Introduction

Malaria is a major public health issue in Cameroon, particularly in rural and semi-urban areas. Accurate prediction of malaria outbreaks can help health authorities allocate resources more effectively, raise awareness, and take preventative actions. This project applies data mining techniques to historical health, weather, and demographic data to build predictive models for malaria outbreaks.

### 2. Project Objectives

* To identify patterns and factors that contribute to malaria outbreaks in South West regions of Cameroon.
* To develop predictive models that can forecast malaria incidence based on weather and environmental conditions.
* To support public health planning with data-driven insights.

### 3. Project Phases

#### • Phase 1: Data Collection

Collect data relevant to malaria cases and influencing factors.

**Sources:**

* Ministry of Public Health reports (malaria case records)
* National Meteorological Center (weather data)
* WHO malaria reports
* NGO reports and surveys

#### • Phase 2: Data Preprocessing

**Data Cleaning:**

* Handle missing values in health and weather datasets.
* Remove duplicate records.
* Correct inconsistencies (e.g., standardize date formats, locations).

**Data Transformation:**

* Normalize temperature, humidity, and rainfall values.
* Convert categorical data into numerical format using encoding.
* Aggregate daily data into weekly or monthly summaries for modeling.

#### • Phase 3: Exploratory Data Analysis (EDA)

**Statistical Analysis:**

* Analyze malaria case trends by region and time.
* Correlate environmental variables with malaria incidence.

**Visualization:**

* Line charts for trend analysis
* Heatmaps for correlation
* Geographic maps to identify high-risk areas

#### • Phase 4: Modeling Techniques

**Classification:**

* Predict whether a month/region is likely to experience a malaria outbreak (yes/no) using Decision Trees or Random Forests.

**Clustering:**

* Segment regions based on similar malaria patterns and weather conditions using K-Means.

**Association Rule Learning:**

* Discover rules such as: "If rainfall is high and humidity > 70%, then risk of malaria outbreak increases."

**Regression:**

* Predict the number of malaria cases in a given region and month using Linear Regression or Random Forest Regressor.

#### • Phase 5: Model Evaluation

**Performance Metrics:**

* Accuracy, Precision, Recall, F1 Score (for classification)
* RMSE, MAE (for regression)

**Validation Techniques:**

* Train-test split
* K-fold cross-validation
* Confusion matrix analysis

#### • Phase 6: Deployment

* Develop a simple web dashboard for health authorities to input current weather and location data and receive outbreak predictions.
* Integrate with SMS/email alert systems for local health centers.

### 4. Tools and Technologies

* **Languages:** Python (pandas, scikit-learn, matplotlib, seaborn)
* **Tools:** Jupyter Notebook, Power BI or Tableau (for visualization), Flask (for deployment)
* **Databases:** MySQL or MongoDB
* **Platforms:** GitHub, Google Colab

### 5. Conclusion

This project has the potential to contribute significantly to the fight against malaria in Cameroon. By leveraging data mining techniques, health authorities can make proactive and informed decisions, ultimately saving lives and improving healthcare delivery