

Jenkins Antimalarial Dataset Documentation

Introduction

Malaria remains a significant public health challenge, particularly in Africa, where it significantly impacts child mortality. Nearly one child under the age of five dies every minute from malaria in Africa alone. Accurate forecasting of antimalarial drug needs is crucial for effective healthcare planning and resource allocation.

The goal of this project is to develop a robust predictive model for forecasting the demand for antimalarial drugs in Africa. By leveraging AI technology, we aim to model complex variables that affect drug demand, such as seasonal outbreaks, population growth, and improvements in healthcare access. This project seeks to contribute to better healthcare management and ultimately reduce the burden of malaria in Africa.

Metadata

- **Title** : Jenkins Anti-Malaria Dataset
- **Description** : This dataset contains information related to anti-malaria interventions and their impact across various countries, including socio-economic and demographic details.
- **Source** : Artificial intelligence(AI) technologies, United Nations Children's Fund (UNICEF)
- **Date** : May 23rd,2024

Data Dictionary

Column Name	Description	Data Type	Example Value
Subdomain	Specific subdomain of anti-malaria intervention	String	"IPTP"
ISO	ISO Country code	String	"AGO"

Countries and areas	Name of country or area	String	"Angola"
Key	Unique identifier for the country-year record	String	"AGO-2016"
UNICEF Reporting Region	UNICEF reporting region name	String	"Eastern and Southern Africa"
UNICEF Programme Region	UNICEF programme region code	String	"ESARO"
World Bank Income Group (2022)	World Bank classification of the country's income group for the year 2022	String	"Lower middle income"
Year	Year of data collection	integer	2016
Population_growth(%)	Annual population growth rate (%)	Float	3.586211
% of children under age 5 with fever	Percentage of children under age 5 reported to have had fever in the past two weeks	Float	18.1
Rural	Percentage of the population in rural areas	Float	11.3
Urban	Percentage of the population in urban area	Float	24.0
Poorest	Percentage of the population in the poorest wealth quintile	Float	8.3
Second	Percentage of the population in the second wealth quintile	Float	15.4
Middle	Percentage of the population in the middle wealth quintile	Float	21.9

Fourth	Percentage of the population in the fourth wealth quintile	Float	23.3
Richest	Percentage of the population in the richest wealth quintile	Float	31.3
None	Percentage of the population with no education	Float	13.1
Primary	Percentage of the population with primary education	Float	17.0
Sec & Higher	Percentage of the population with secondary or higher education	Float	26.2

Data Format

- File Type: Excel (.xlsx)
- Encoding: UTF-8

Usage Information

- Licensing: The dataset is available under the Creative Commons Attribution 4.0 International License (CC BY 4.0)
- Citations: Please cite as "Jenkins Research Lab, Jenkins Anti-Malaria Dataset, 2024".

* Example usage

```
import pandas as pd

# Load the dataset from the provided Excel file
file_path = '/content/Jenkins_AntiMalaria_Dataset.xlsx'
data = pd.read_excel(file_path)
# Display the first few rows
print(df.head())
```

Quality and Limitations

- Data Quality: The dataset has been carefully curated; however, some socio-economic data may have estimation errors.
- Limitations: Data may not cover all regions equally, leading to potential biases. Some countries or areas might have incomplete data for certain years.

Preprocessing Steps

- Data Cleaning: Missing values in demographic and socio-economic indicators were filled using appropriate statistical methods. Ensured consistency in country names and codes.
- Transformations: Created new features such as Education_Level by combining columns Primary and Sec & Higher. Standardized percentage values for comparison across different metrics.