FTL Global ML bootcamp 3 | Capstone project Idea proposal | Group 13

Malnutrition Risk Prediction ML Model

Group members:

- Fatima Nasraldin
- Lina Ahmed
- Linda Adil
- Maha Abdalfedil
- Nada Ali

1. Project Idea:

This project aims to develop a machine learning model for predicting malnutrition risk trends over time by leveraging historical health, nutrition, and population data.

The model will classify children into high-risk, medium-risk, or low-risk categories for malnutrition, based on a comprehensive set of features including socio-economic conditions, dietary intake, household demographics, health indicators, and environmental factors.

This classification will enable targeted interventions and inform decision-making for addressing malnutrition effectively.

2. Relevance to Sustainable Development Goals (SDGs):

The project supports *SDG 2* (***Zero Hunger***) by addressing malnutrition, a key aspect of global hunger, through predictive models that guide targeted interventions in vulnerable areas.

It also aligns with *SDG 3* (***Good Health and Well-being***) by enabling better management of nutritional needs, leading to improved health outcomes.

Additionally, the project contributes to *SDG 17* (***Partnerships for the Goals***) by offering data-driven insights to facilitate collaboration between governments, NGOs, and international organizations for more effective nutrition interventions

3. Literature Examples:

- *Forecasting Child Malnutrition in Developing Countries: Evidence from Multiple Data Sources* Reference
- *Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia* *Reference*

4. Data Description:

The data for this project is sourced from the World Bank Group's Health [*Click here to access the data*] , Nutrition, and Population dataset, available in CSV format. It includes multiple years of data with various health and nutritional indicators across several rows and columns.

Preprocessing will involve handling missing values and outliers, resampling for consistent time intervals, and engineering features like rolling averages and lagged variables to capture temporal dependencies

5. Approach:

The project will use ***machine learning*** techniques due to:

the nature of the dataset, which consists of ***time series data*** with relatively simple features. This makes models like Random Forests, Gradient Boosting, and CatBoost well-suited for capturing patterns and trends effectively.

Justification:

Machine learning models are appropriate for ***forecasting*** tasks with structured time series data, allowing for effective trend analysis and risk prediction without the need for deep learning's computationally intensive techniques