AI Development Workflow Assignment - Part 1

1. Problem Definition

Hypothetical AI Problem:

"Predicting Malnutrition Risk Among Children Under 5 in Rural Kenya."

Objectives:

1. Identify children at high risk of malnutrition using health, socioeconomic, and geographic data.

2. Enable early intervention by community health workers.

3. Inform policy decisions on food aid and healthcare allocation.

Stakeholders:

- Ministry of Health, Kenya.

- UNICEF and NGOs focused on child welfare.

Key Performance Indicator (KPI):

- F1 Score of the predictive model in identifying high-risk children.

2. Data Collection and Preprocessing

Two Data Sources:

1. Demographic and Health Surveys (DHS) Kenya - Contains child anthropometry and household characteristics.

2. Kenya Integrated Household Budget Survey (KIHBS) - Includes economic, education, and food security indicators.

One Potential Bias:

- Urban bias - Surveys may overrepresent urban populations, underestimating rural-specific risk factors.

Three Preprocessing Steps:

1. Handling Missing Data: Use imputation methods (mean, median, or regression-based).

2. Normalization: Scale continuous variables to bring features to a similar range.

3. Encoding Categorical Features: Apply one-hot encoding to variables like "region" or "maternal education level."

3. Model Development

AI Development Workflow Assignment - Part 1

Chosen Model:

- Random Forest Classifier

Justification:

- Performs well on tabular data, handles non-linear relationships, is robust to outliers, and provides feature importance

for interpretability.

Data Split:

- 70% Training

- 15% Validation

- 15% Test

Two Hyperparameters to Tune:

1. n_estimators: Number of trees in the forest - affects performance and overfitting.

2. max_depth: Maximum depth of the trees - controls model complexity and generalization.

4. Evaluation and Deployment

Two Evaluation Metrics:

1. F1 Score - Balances precision and recall, crucial when false negatives (undetected malnutrition) are dangerous.

2. ROC-AUC - Measures the ability of the model to distinguish between classes, useful for threshold tuning.

What is Concept Drift?

- Concept drift occurs when the statistical properties of target variables change over time, degrading model accuracy.

- Monitoring Approach: Use real-time dashboards comparing model predictions to actual outcomes; retrain the model

quarterly.

One Deployment Challenge:

- Scalability - Deploying the model across remote rural clinics with limited internet and computing infrastructure.