AGRICULTURAL YIELD PREDICTION SYSTEM FOR FARMERS IN KENYA

CARNEGIE MELLON UNIVERSITY

COURSE: ARTIFICIAL INTELLIGENCE SYSTEM DESIGN

TEAM 1: Ethel Kwagalakwe | Martin Mulang' | Bruce Mugizi | Emrys Nicholaus

JANUARY 25, 2026

PROBLEM STATEMENT

As of 2023, 282 million Africans face hunger daily, a situation that is made worse by climate change. Projection indicates that by 2050, Africa might fulfil only 13% of its required food [1, 2]. The continent's smallholder farmers, who produce up to 80% of its food, are vulnerable to this crisis due to disconnected value chains and heavy reliance on rain-fed agriculture, which accounts for over 95% of farmland [3]. Climate change has brought uncertainties to the weather, making rainfall patterns unpredictable.[4, 3] The high variability of rainfall and moisture leads to low, unstable yields, a problem for food security. Farmers often lack timely, actionable information to adapt, and they require knowledge about optimal planting times and other practices to be able to manage unpredictability. Empowering farmers with actionable advice ahead of time is likely to drive up production and slow the decline in crop yields [5]. Our project aims to develop an AI-powered advisory system that provides planting guidance. The system uses historical weather and yields data to model and predict future rainfall and temperature patterns. It performs classification analysis on historical yield outcomes to recommend optimal practices. Based on these forecasts and classifications, the system employs automated logic to generate advisory messages, which are delivered directly to farmers via SMS.

DATASETS AND SOURCES

This system will use meteorological and agriculture datasets to train models for rainfall forecasting and yield risk classification.

Rainfall Data (TAMSAT): Link: https://research.reading.ac.uk/tamsat/data-access/

This repository provides high-resolution, satellite-derived estimates of rainfall for Africa. Data is stored in NetCDF format, containing daily rainfall grids over the continent. For model training, we will extract historical rainfall time series for specific districts using their latitude and longitude coordinates.

Agricultural Yield Data (FAO & GROW-Africa): Link: https://data.apps.fao.org/

We will use historical yield observations for target crops (e.g., maize, cassava) at a sub-national or district level. This data will be used to calculate statistics (mean yield, standard deviation) and create a labeled dataset for classification (e.g., "Low," "Medium," "High" yield risk).

MACHINE LEARNING ARCHITECTURE AND AI TASK

Rather than relying on a single model, we will use a multi-layered approach to ensure both accuracy and practical utility:

The system employs a multi-layered engine to translate weather data into planting advice. Random Forest and LightGBM identify key variables, such as moisture and temperature, that affect crop success. LSTM networks model sequential patterns in TAMSAT rainfall data, while a hybrid TabNet approach integrates historical yields via late fusion. SHAP provides explainability, ensuring recommendations are transparent and trustworthy for farmers.

Deployment Target: Cloud-based (Google Colab, AWS/GCP for production) with potential edge deployment demonstration on Raspberry Pi for resource-constrained educational institutions.

SUCCESS METRICS AND EVALUATION

System performance is assessed through predictive accuracy and classification reliability. Rainfall forecasts are evaluated using Root Mean Square Error to penalize significant deviations. Yield risk classification is measured by Precision, Recall, and F1-Score, prioritizing high recall to identify all high-risk periods. The Disparate Impact Ratio is monitored to ensure equitable, unbiased advice across all regions and demographics.

PREDICTED CHALLENGES AND MITIGATION STRATEGIES

Technical solutions address key data and deployment hurdles. SMOTE oversampling and focal loss mitigate class imbalance caused by rare yield failures. Attention mechanisms manage variable rainfall sequences by focusing on critical germination phases. To ensure accessibility, we use model quantization to keep the system under 100MB, achieving the low latency required for SMS delivery and edge devices.

RESOURCES AND IMPLEMENTATION TIMELINE

The project will employ a tiered, cost-effective computational strategy. Model development and training will primarily utilize Google Colab's free T4 GPU resources, with Kaggle / Colab Pro as a backup. For deployment, a lightweight API will be hosted on cloud free tiers: AWS/GCP, while a Raspberry Pi 4 will demonstrate edge feasibility. The core software stack is Python-based, using PyTorch/TensorFlow for AI modeling, Scikit-learn for traditional ML, and essential data libraries such as Pandas and NumPy. SHAP will provide model explainability, and FastAPI/Streamlit will power the deployment interface and demo dashboard.

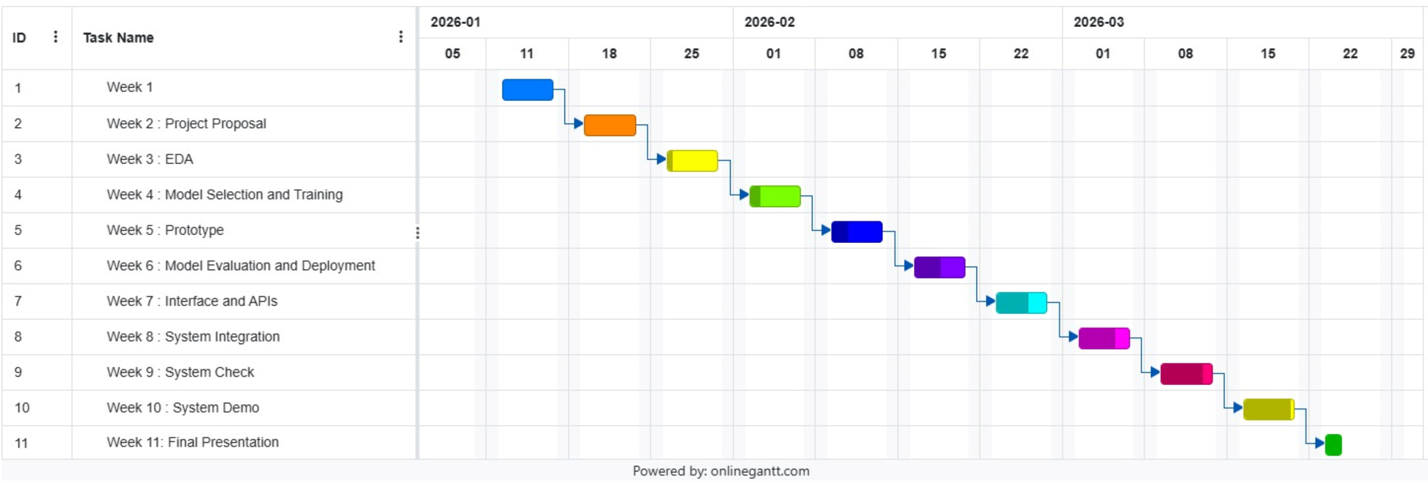

Figure 1: Project timeline

EXPECTED IMPACT AND SDG ALIGNMENT

This project provides a deployable forecasting system for integration into mobile platforms like mAgri or iCow. By delivering yield warnings up to three months in advance, the system empowers farmers to adjust irrigation and planting schedules while enabling governments to pre-position food aid. These forecasts also support index-based crop insurance for rapid payouts and help commodity traders stabilize supply chains. Research indicates that such accurate modeling can reduce post-harvest losses by up to 20% and improve input efficiency by 30%. For smallholders, this intervention directly strengthens food security and builds essential climate resilience in line with Sustainable Development Goals.

REFERENCES

[1] John Choptiany. Africa’s smallholder farmers face collapse if we do not act on climate change. https://www.cgiar.org/news-events/news/ africas-smallholder-farmers-face-collapse-if-we-do-not-act-on-climate-change, August 2025. Accessed: 2026-01-25.2

[2] Brenda Coromina. Amidst growing population pressures, who will feed Africa? https://www. ilri.org/news/amidst growing-population-pressures-who-will-feed-africa, 2021. Accessed: 2026-01-25.

[3] Digital Earth Africa. Creating an open-source framework for crop-type mapping in africa. https://digitalearthafrica.org/en_za/ creating-an-open-source-framework-for-crop-type-mapping-in-africa/, June 2024. Accessed: 2026-01-25.

[4] Tshepo S. Masipa. The impact of climate change on food security in south africa: Current realities and challenges ahead. J`amb´a: Journal of Disaster Risk Studies, 9(1):411, 2017. PMCID: PMC6014268; Accessed: 2026-01-25.

[5] Disha Shetty. Food insecurity affects 282 million people in 2023. https:// healthpolicy-watch.news/food-insecurity affects-282-million-people-in-2023/, 2024. Accessed: 2026-01-25.