# 

# Forecasting Farmer Futures



Idea Proposal

**Group 4**

**Kyaw Phyo Aung**

**Aung Htet Paing**

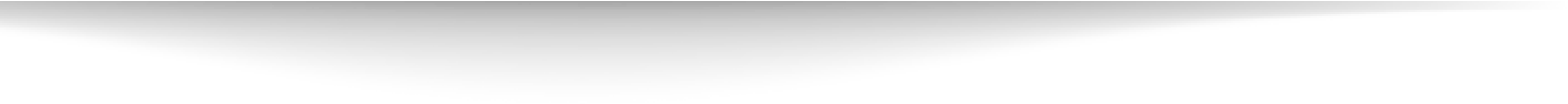
**Khin Chaw Lai Lai Tun**

**Thaw Maung Oo**

**Saw Htet Khine Soe**

**Yu Thazin**

**Prepared for**



FTL Myanmar Machine Learning Bootcamp

### Project Idea

Myanmar's smallholder farmers face significant risks from climate change and economic instability, leading to unpredictable yields and income loss. We propose building a predictive AI engine that analyzes historical climate, economic, and agricultural data to forecast crop yield and profitability. The goal is to provide a data-driven decision-support tool for farmers to select the most resilient and profitable crops, enhancing their livelihoods and food security.

### 2. Relevance to Sustainable Development Goals (SDGs)

This project directly supports key SDGs by enhancing agricultural resilience:

* SDG 1 (No Poverty): Stabilizes and increases farmer incomes by minimizing crop failure.
* SDG 2 (Zero Hunger): Improves food security by optimizing crop selection for better yields.
* SDG 13 (Climate Action): Provides a practical adaptation tool for farmers to build resilience against climate impacts.

### 3. Literature Examples

Our approach is validated by existing academic research:

1.["**Case study on climate change effects and food security in Southeast Asia**"](https://www.nature.com/articles/s41598-024-65140-y) (Taniushkina, et al., 2024): This paper validates using a multi-source data approach for crop prediction in Southeast Asia (including Myanmar) and confirms XGBoost as a top-performing model, directly supporting our chosen methodology.

2.["**A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches**](https://www.researchgate.net/publication/351116727_A_Comprehensive_Review_of_Crop_Yield_Prediction_Using_Machine_Learning_Approaches_With_Special_Emphasis_on_Palm_Oil_Yield_Prediction)" (Rashid, et al., 2021): This review provides a foundational overview of effective ML models (ANN, RF) and key features (climate, soil), which will guide our model and feature selection process

### 4. Describe Your Data

We will build a time-series dataset (2010-2025) by integrating multiple open-source data:

* Sources: MIMU (Agriculture & Economy), Government Reports (Soil), World Bank/NASA (Climate), and local sources (Market Prices).
* Features: Key variables will include crop production statistics, soil types, temperature, precipitation, and input costs (e.g., fertilizer, fuel).
* Format: The raw data (CSV, Geospatial) will be collected, cleaned, and preprocessed into a unified dataset ready for model training.

### 5. Approach (Machine Learning or Deep Learning)

Our core strategy is to use XGBoost (Extreme Gradient Boosting), a powerful and widely-used machine learning model. We selected XGBoost for its high predictive accuracy on structured datasets like ours and its crucial feature of interpretability. Unlike a "black-box" model, XGBoost can generate "feature importance" scores, allowing us to identify and clearly explain which factors—such as rainfall, temperature, or fertilizer prices—have the most significant impact on agricultural outcomes. This explainability is essential for providing actionable insights to farmers and stakeholders.

As a potential extension for more advanced forecasting, we will explore a Long Short-Term Memory (LSTM) neural network. Given that our data spans many years, an LSTM is uniquely suited to capture complex time-series patterns and long-term dependencies that simpler models might miss. This could enhance the model's ability to make more nuanced long-range predictions.