**3. Project Workflow**

**Phase 1: Data Preprocessing**

1. **Data Cleaning**:
   * Handle missing values or outliers in phenotypic and environmental data.
   * Normalize/scale the data for deep learning models.
2. **Exploratory Data Analysis (EDA)**:
   * Visualize relationships between plant traits, temperature, chlorophyll levels, and yield (KW).
   * Identify correlations between environmental stress (e.g., canopy temperature) and plant performance.

**Phase 2: Data Augmentation with GANs**

1. Implement a **Conditional GAN**:
   * Input: A treatment type (e.g., T1, T2) and random noise.
   * Output: Synthetic data for phenotypic and environmental attributes.
2. Train the GAN on the existing data.
3. **Validate the synthetic data**:
   * Check if generated data follows similar statistical properties as the real data.
4. Add augmented data to the original dataset.

**Phase 3: Deep Learning for Yield Prediction**

1. **Model Development**:
   * Use a neural network model (e.g., **MLP**, **LSTM**, or **CNN**) to predict **kernel weight (KW)**.
   * Input: All phenotypic traits, canopy temperature, and chlorophyll data.
   * Output: Kernel weight (KW) as the target variable.
2. **Model Training and Validation**:
   * Train the model on the combined (real + GAN-generated) dataset.
   * Use **cross-validation** to ensure robust performance.
3. **Evaluation**:
   * Use metrics like **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, and **R² score**.

**Phase 4: Model Interpretability with XAI**

1. Use **SHAP** or **LIME** to analyze feature importance:
   * Identify which features (e.g., canopy temperature, plant height, chlorophyll content) contribute most to kernel weight.
2. Visualize results to show how different traits impact yield.

**Phase 5: Stress Analysis and Insights**

1. Analyze relationships between:
   * High canopy temperature and reduced chlorophyll levels.
   * Yield reduction under stress conditions.
2. Generate insights:
   * Recommend ideal plant traits for higher yields.
   * Suggest interventions for stress-tolerant genotypes.

**4. Project Flowchart**

Here’s the visual flow of the project:

1. **Data Collection**  
   ⬇
2. **Data Preprocessing & EDA**  
   ⬇
3. **Data Augmentation (GANs)**  
   ⬇
4. **Deep Learning (Yield Prediction)**  
   ⬇
5. **Explainable AI (SHAP/LIME)**  
   ⬇
6. **Stress Analysis & Insights**  
   ⬇
7. **Recommendations for Yield Optimization**

**5. Predictions Using the Dataset**

The main prediction task is:

* **Kernel Weight (KW)**: Predict maize yield using plant phenotypic and environmental traits.

Additional insights:

1. **Stress Detection**:
   * Predict yield decline under high canopy temperature or low chlorophyll levels.
2. **Genotype Recommendations**:
   * Identify the most suitable genotypes (T1–T8) for optimal yield based on environmental conditions.

**6. Tools and Libraries**

* **Data Preprocessing & EDA**: Python, Pandas, Numpy, Matplotlib, Seaborn.
* **GANs**: TensorFlow/Keras or PyTorch.
* **Deep Learning**: TensorFlow/Keras or PyTorch for model development.
* **XAI**: SHAP, LIME libraries for explainability.
* **Model Evaluation**: Scikit-learn for metrics like MAE, MSE, and R².