**Digital Green Crop Yield Estimate Challenge**

1. **Data Acquisition:** 
   1. Obtained data (Train.csv & Test.csv) from the Challenge Repo
2. **Exploratory Data Analysis (Data Understanding & Preparation):**
   1. Load Dataset & analyze few random samples
   2. Obtain all features from the data
   3. Display data type information of each attribute & differentiate categorical, continuous features
   4. Identify missing values and impute them in both categorical and continuous features
   5. Encode Categorical Features
   6. Perform Outlier Analysis & display box plots for continuous features
   7. Display Data Quality Report for both categorical and continuous features
   8. Display Correlation Matrix for both categorical and continuous features
   9. Plot Histograms for Continuous features and Barcharts for Categorical features
3. **Modeling:**
   1. Train Linear Regression, Support Vector Regressor, Random Forest Regressor, XGBoost Regressor, AdaBoost Regression on dataset obtained after initial EDA
   2. Perform Normalization on EDA file & Repeat step 3.1. If Normalization improves the performance, use Normalized features throughout.
   3. Apply Dimensionality Reduction Techniques (PCA, ICA, LDA, tSNE) and & Repeat step 3.1. Select the best performing dimensionality reduction technique and use throughout.
   4. Design a Regularized DNN for Regression and train on cleaned dataset
   5. Design a Regularized DNN with Attention for Regression and train on cleaned dataset
   6. Design a LSTM for Regression and train on processed dataset (Will this work??)
   7. 1D CNN for Regression??
4. **Evaluation:**
   1. The evaluation metric for this competition is [Root Mean Squared Error](https://zindi.africa/learn/zindi-error-metric-series-what-is-root-mean-square-error-rmse#:~:text=Root%20Mean%20Squared%20Error%20or,want%20the%20model%20to%20achieve.). Validate on 20% of hold-out dataset from Train.csv
5. **Deployment / Submission:** 
   1. Make submission on Test.csv predictions