**Crop Production Classification Documentation**

**1. Import Libraries:** Essential libraries like pandas, NumPy, Matplotlib, Seaborn, and ScikitLearn are imported to handle data processing, visualization, and machine learning.

**2. Load Data:** The dataset is loaded using pandas, providing a foundation for subsequent analysis.

**3. Exploratory Data Analysis (EDA) Before Feature Engineering:** Calculate statistics (mean, median, mode) for numerical features. Visualize distributions using histograms and boxplots. Analyze relationships between features and target using scatter plots and correlation matrices. Examine categorical variable distributions through bar charts.Identify missing values and detect outliers.

**4. Feature Engineering:** Create lagged variables or rolling averages for timeseries data.Transform skewed variables using log transformations. Generate new features or bin continuous variables. Drop or adjust features with low relevance.

**5. EDA After Feature Engineering:** Recheck the distribution of engineered features. Visualize relationships between new features and target. Compare pre and posttransformation distributions.

**6. Data Preprocessing:** Handle missing data using `SimpleImputer`. Encode categorical features with `OneHotEncoder`. Scale numerical features using `StandardScaler`.

**7. Data Splitting:** Divide data into training and testing sets using `train\_test\_split`.

**8. Model Definition and Pipelines:** Create pipelines for models like Random Forest, Logistic Regression, Decision Tree, and others. Each pipeline includes preprocessing and model fitting.

**9. Training and Evaluation:** Train models, make predictions, and evaluate using accuracy, precision, recall, F1 score, ROCAUC, and confusion matrices.

**10. CrossValidation and Hyperparameter Tuning:** Use `cross\_val\_score` for model evaluation. Apply `GridSearchCV` to optimize model parameters.

**11. Voting Classifier & Evaluation:** Combine models into a `VotingClassifier`. Evaluate ensemble performance through a classification report. Plot the ROC curve, calculate AUC, and visualize the model’s discriminative power.

**Novelty: -** We created a new feature ‘production\_trend\_1995\_2013’ which helps us to classify the production trend such as “Upper Trend, Lower Trend” over the years in different areas and different crops. We also created new feature as “production\_lag”.