Project 4: Prediction of Agriculture Crop Production in India

Company: UCT (UniConverge Technologies Pvt. Ltd.)
Internship Domain: Machine Learning
Duration: 6 Weeks
Project Domain: Agriculture

1. Background
India, with a population of over 1.3 billion, is the second-largest country by population and one of the world's largest agricultural economies. A significant portion of the Indian population depends on agriculture for their livelihood. Predicting crop production accurately is critical to ensuring food security, planning logistics, and determining market policies.

This project, undertaken as part of the UCT Machine Learning Internship, uses data from data.gov.in to build a machine learning model for predicting crop production.

2. Problem Statement
Given crop-specific data including cost of cultivation, yield, and crop type, predict the annual production for a given crop using machine learning.

3. Relevance
This project has strong practical significance for government, farmers, and agribusinesses. An accurate production prediction model can help in planning storage, pricing, procurement, and subsidies.

4. Design and Approach
The project involved the following steps:
- Merging and cleaning multiple crop-related datasets from government sources
- Handling missing data using mean/mode imputation
- Encoding categorical variables such as crop, zone, and season
- Feature selection based on availability and impact

- Model training using Random Forest Regressor
- Performance evaluation using RMSE and R² score

5. Implementation
We used Python with Pandas and scikit-learn for preprocessing and model building. The final model was trained using features like Crop, Year, Yield, and Cost per Quintal. One-hot encoding was used for categorical data, and missing numeric values were filled using mean imputation.

6. Results
The Random Forest model was trained successfully and evaluated using:
- Root Mean Square Error (RMSE)
- R² Score (Goodness of Fit)
The model produced acceptable error margins, making it useful for large-scale estimation.

7. Learnings
From this project, I learned how to:
- Merge diverse real-world datasets
- Clean and impute missing agricultural data
- Encode categorical data for ML models
- Train and evaluate regression models
- Use agriculture domain knowledge to select relevant features