Crop Yield Prediction and Profit Optimisation

Harsh Sinha (IMT2023571)
Kartikeya Dimri (IMT2023126)
Syed Naveed (IMT2023119)

Agriculture: The Backbone of India

* Statistically a significant majority of India earns their livelihood through farming.
* Our project aims to enhance and maximise the income of this demography.

Challenges faced by farmers

HARVEST REWARDS

Crop	New MSP*	% (Absolute*) hike over 2023-24 (In ₹)
Niger seed	8,717	12.7% (983)
Ragi	4,290	11.5% (444)
Tur/Arhar	7,550	7.8% (550)
Sunflower seed	7,280	7.7% (520)
Cotton	7,121	7.5% (501)
Sesamum	9,267	7.3% (632)
Urad	7,400	6.5% (450)
Groundnut	6,783	6.4% (406)
Maize	2,225	6.4% (135)
Soybean	4,892	6.3% (292)
Jowar	3,371	6% (191)
Paddy	2,300	5.3% (117)
Bajra	2,625	5% (125)

*per quintal **Paddy** (Common); **Jowar** (Hybrid); **Cotton** (Medium Staple)

Problem Statement

* Predict crop yield in a specific state using regression models.
* Estimate profitability by integrating predicted yield with predicted crop prices (using historical MSP values).
* Optimise crop selection and allocation using linear programming so that farmers maximise profit under resource constraints.

Part I: Predicting Yield and MSP

* Using the data of previous year crop-wise yield, we are formulating a regression problem to predict the yield for next year.
* Using the historical MSP data provided on various crops by the Government, we will try to predict the MSP for the upcoming year.

Dataset 1: Crop Yield

Crop Name	Crop Year	Season	State	Area (ha)	Production (ton)	Annual Rainfall (mm)	Fertilisers (kg)	Pesticides (kg)	Yield (ton/ha)
Arecanut	1997	Whole Year	Karnataka	93100	133342	1266.7	8860327	28861	1.29

* Total Records: 19690
* Data from 1997 to 2019
* Source: Kaggle

Dataset II: MSP

Year	Crop Name	MSP (INR per quintal)
1998-1999	Gram	Rs. 895

* Total Records: 760
* Data from 1998 to 2025
* Source: Kaggle & Government MSP statement

Easy Regression

* Ordinary Least Squares (OLS): OLS minimises the sum of squared residuals

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - X_i\beta)^2$$

Easy Regression

* Polynomial Regression An extension of OLS by adding polynomial terms of features:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_d x^d + \epsilon$$

Advance Regression

* Ridge Regression (L2 Regularization)

$$\hat{\beta} = \arg\min_{\beta} \left(\sum_{i=1}^{n} (y_i - X_i\beta)^2 + \lambda \|\beta\|_2^2 \right)$$

* Lasso Regression (L1 Regularization)

$$\hat{\beta} = \arg\min_{\beta} \left(\sum_{i=1}^{n} (y_i - X_i\beta)^2 + \lambda \|\beta\|_1 \right)$$

* Elastic Net

$$\hat{\beta} = \arg\min_{\beta} \left(\sum_{i=1}^{n} (y_i - X_i\beta)^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \right)$$

Advance Regression

* XGBoost Regression (Linear Booster) XGBoost optimizes a regularized objective function:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{k} \Omega(f_k)$$

where l is the loss (squared error for regression) and Ω(fk) is the regularization term:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|\omega\|^2$$

* Here T is the number of leaves in the tree, ω are leaf weights, and λ, γ are regularization parameters. In linear booster mode, XGBoost reduces to a highly optimized regularized OLS solver.

Part II: Optimisation Using Linear Programming

* After predicting yield and price, we solve an optimisation problem. Linear Programming (LP) is the mathematical method used to solve this optimisation problem.

* It is perfectly suited for this task because both the objective function (total profit) and the constraints (land, resources) are linear relationships.

Reference: http://cgm.cs.mcgill.ca/~avis/courses/567/notes/stoch1.pdf

Objective Function

$$\text{Profit} = \sum_i \Big(\text{PredictedYield}_i \times \text{PredictedPrice}_i \times \text{LandAllocated}_i - (\text{FertilizerCost}_i + \text{PesticideCost}_i + \text{MiscCost}_i) \Big)$$

where i = crop index and costs include fertilizer, pesticide, and optional miscellaneous inputs given by the farmer. Here, the decision variable that the model needs to determine is land allocated for each crop.

Constraints

* Total Land Constraint:

$$\sum_i \text{LandAllocated}_i \leq \text{Farmer's Land}$$

* Fertiliser Constraint:

$$\sum_i \text{FertilizerUsage}_i \leq \text{Available Fertilizer}$$

* Pesticide Constraint:

$$\sum_i \text{PesticideUsage}_i \leq \text{Available Pesticide}$$

* Seasonal Constraint: Each crop can only be allocated to its valid season (Rabi, Kharif, Whole Year).

Real-Life Application

* Farmer inputs: state, land size, available fertilizers/pesticides, input prices (fertilizers, pesticides, miscellaneous).
* Model predicts yields for all crops possible in that state.
* Price model predicts MSP-based crop prices.
* Linear programming optimizes crop allocation to maximize total profit.
* Output: Recommendation for Rabi, Kharif, and Whole-year crops with expected profit.

Impact

* For Farmers: Provides clear, data-driven crop planning strategies, maximizing income.
* For Policymakers: Helps in identifying profitable crops per region, guiding subsidy and MSP policies.
* By integrating machine learning with optimization, this project provides a real-world decision support tool that can directly impact farmer welfare.

Thank You!

FARMERS WELL WISHERS
PRESENTS
ANNADATA SUKHIBHAVA
Progress of the Nation with Country Cow