KHETI - Knowledge-driven Harvest for Ecological Transformation and Innovation

Sachit Vij, Sankalp Tattwadarshi Swain

OBJECTIVE (clearly define what do you want to do)

We want to analyze soil quality, soil fertility of a farm, group the total area into multiple zones and then recommend crops to be grown in those zones which maximize profit and minimize crop loss.

MASTERPLAN (think about how you would implement this and list steps)

1. Ground survey report of the farm/field
    a. Should include micronutrient conc information, location of sample, yield of sample and market_price of target crops
    b. Can include
        i. Water supply information
            1. Underground water levels
            2. Nearest surface water resource
            3. Existing irrigation facilities
        ii. Any annual or biannual crop disease or pest infestation specific to the region

2. Define soil fertility **[DONE]**
    a. It is an index based on yield of the crop produce
    b. This is crop specific
    c. Use historical data

3. Define soil quality **[DONE]**
    a. It is based on organic matter and moisture retention, along with micronutrient concentration values (Macronutrient matters more than micro; indian soils tend to have the same concentration for micro mostly. Focus on NPK)
    b. This is not crop specific; this is a characteristic of the soil
    c. Use multi decision criteria to determine soil quality.

INDEX	High Soil Quality	Low Soil Quality
HIgh Soil Fertility	CASE A	CASE B
Low Soil Fertility	CASE C	CASE D

High Yield		Wheat, Barley
Low Yield		Mustard, Bajra
INDEX	Low SFI	High SFI

4. APPROACH FOR ZONING THE FARM AREA
    a. Approach 1:
        i. We form groups based on micronutrient concentration only (find similar neighbouring points) and then find the most suitable crop for the group
        ii. Similarity between neighbouring points is our concern here
    b. Approach 2:
        i. We pick the best alternative crop and then find the most suitable zone for the crop
        ii. Find all suitable locations for the crop
        iii. Calculate density of suitable location points
        iv. Choose an area within the desired density
        v. Similarity of features of the location point with the ideal point for the crop growth is our concern here

5. Define profit
    a. Linearly optimize the equation

    $$\sum\{(market\_price\ *\ yield\ *\ designated\_area)\ -\ (cost\ of\ input\ for\ crop)\}\ =\ profit$$

    b. Yield is predicted from given data
    Maximise profit given constraints on designated_area (sum of all designated areas = area of farm) {DONE}
    Obtain market_price from survey **{DONE}**
    c. To optimize this further, include extra costs that need to be incurred to grow each crop (more irrigation/particular fertiliser)
    d. Show as benefit is to cost ratio

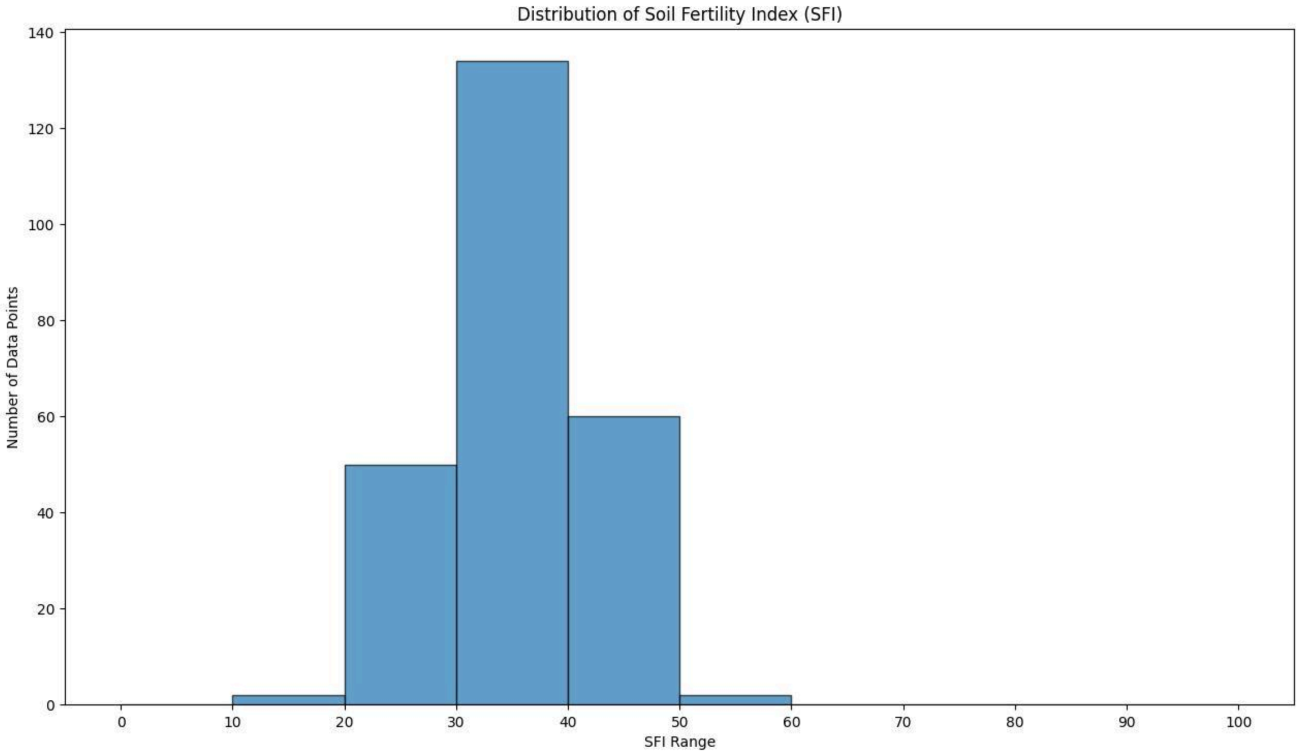
6. Minimize crop loss (TO BE DONE LATER)
    a. Define crop loss first

ISSUES TO BE CONSIDERED

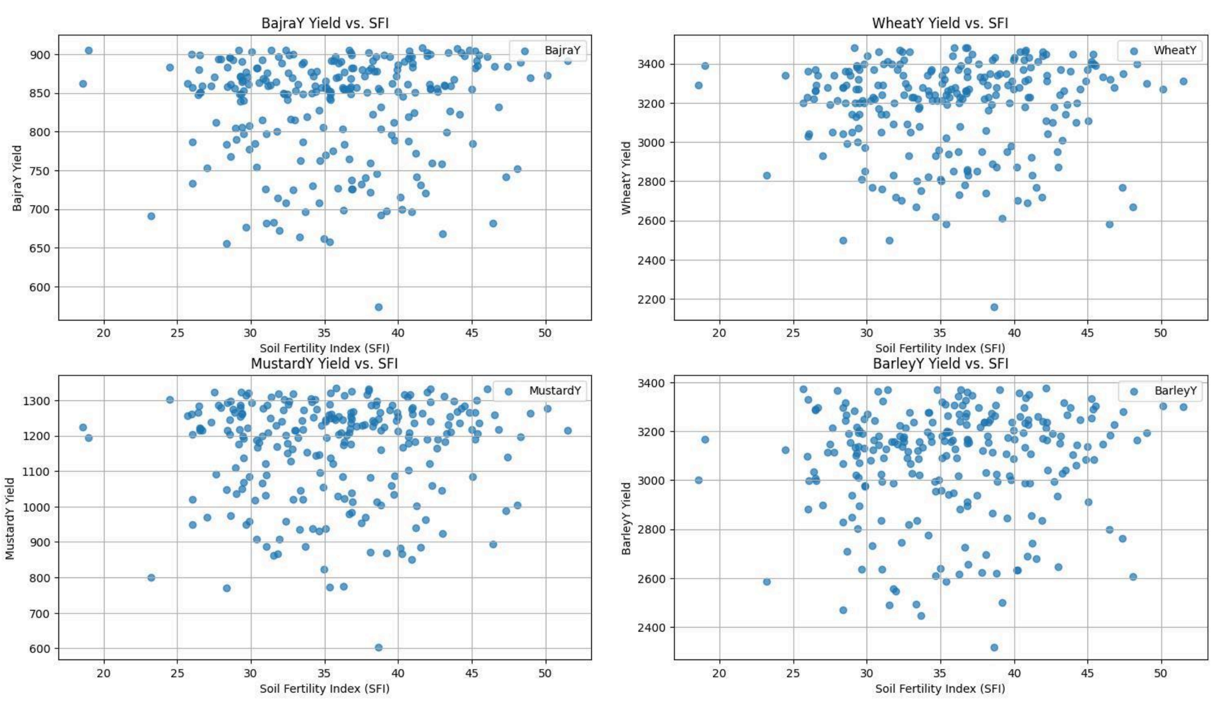
1. The entire region in the experimental dataset is mostly infertile.

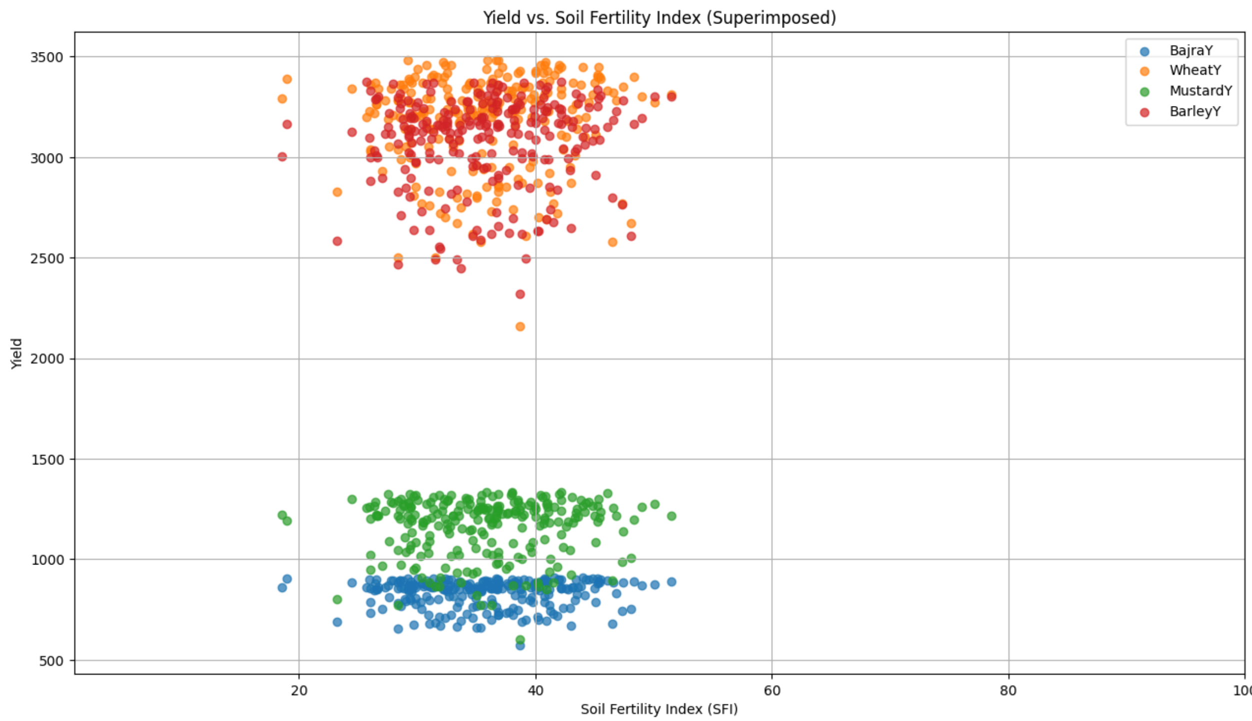
RESULTS AND DISCUSSION

Soil Fertility Index is chosen to represent soil quality, which is purely based on the micronutrient concentration of the soil samples.



As we can notice, the dataset does not include data points in all SFI categories and follows a normal distribution, centred around the range of 30-40 SFI, which indicates that the soil is poor and needs nutrient replenishment.





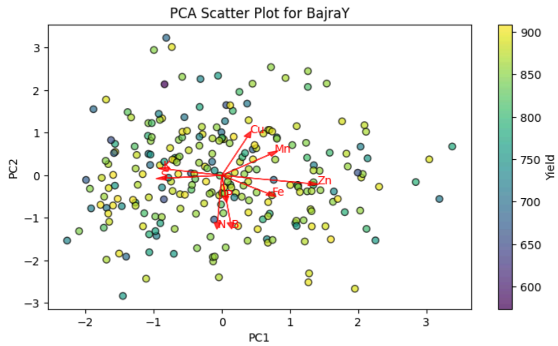
We plotted the yield v/s SFI values for each crop (i) individually and then (ii) in the same plot to visualise the distribution of yield better. We notice these crops have certain belts of concentration with scattered points below them, indicating that these crops yield mostly close to their ceiling values and some are scattered below it. We do not notice any outliers above the threshold.

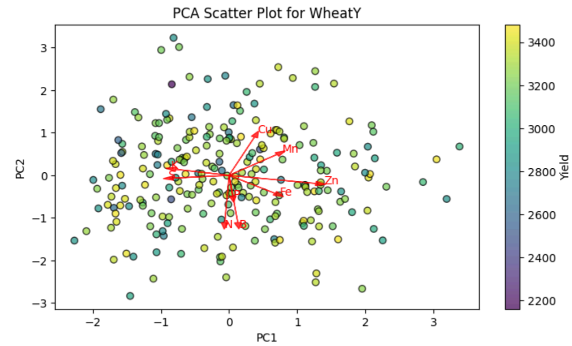
The major issue arises with the dataset. Because for each datapoint in the dataset, the yield values follow the same decreasing order of (wheat > barley > mustard > bajra), the prediction model also outputs a similar result. Considering only Rabi Crops (Wheat, Mustard, Barley), the maximum we can expect from these graphs is to see wheat or barley being recommended based on purely yield.

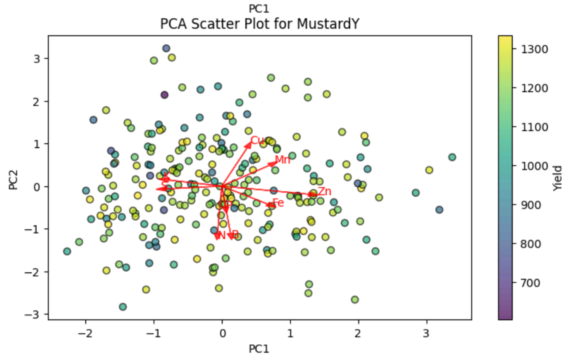
To counter the same, we looked at market prices for these crops and wanted to see if that balances out the disparity in yield. We introduced that bias as a coefficient in the profit equation and based on the MSP values (from the press release by the Ministry of Agriculture & Farmers Welfare, Govt of India). Now we notice a higher margin of profit (MSP - COP) for mustard compared to wheat or barley. However, the profit margin bias dominates the results now and we see mustard as the only recommendation for all areas of growth.

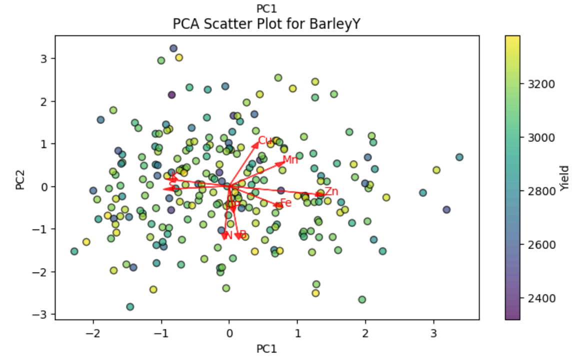
In summary,

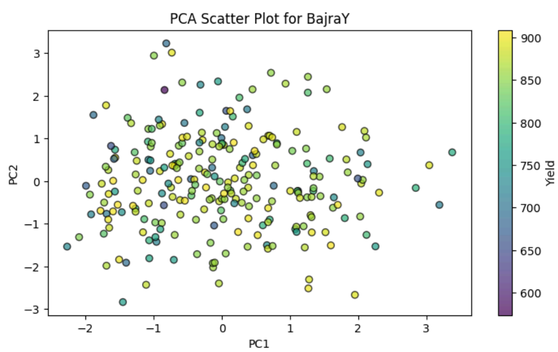
1. There is no datapoint/ SFI value where we notice any crop other than wheat producing higher yield. Hence, only zoning areas based on SFI, does not form a basis for crop recommendation.
2. Bias due to the profit margin of Mustard influences the recommendation drastically. For context, yield is also taken into consideration here.

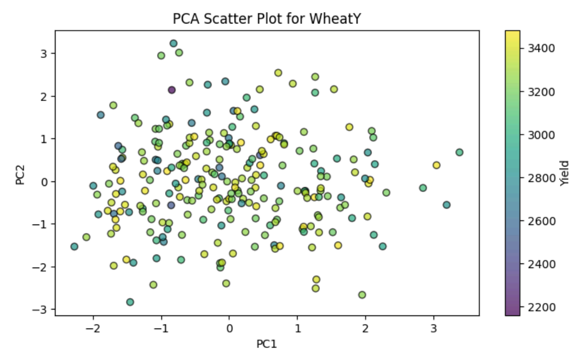

PCA Scatter Plot for BajraY
Cu
Mn
Zn
Fe
N
P
PC1
PC2
Yield

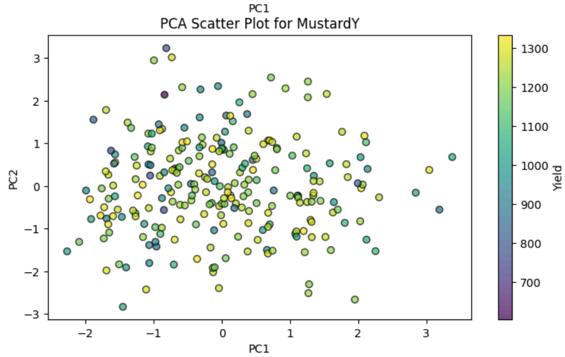

PCA Scatter Plot for WheatY
Cu
Mn
Zn
Fe
N
P
PC1
PC2
Yield

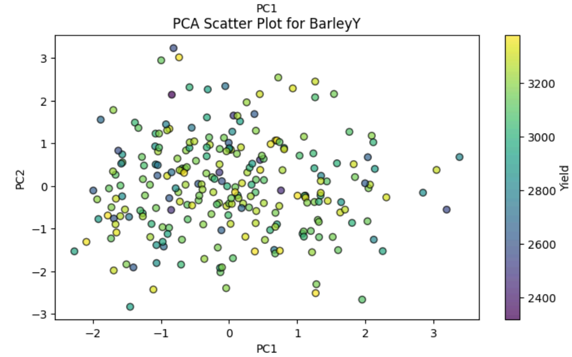

PCA Scatter Plot for MustardY
Cu
Mn
Zn
Fe
N
P
PC1
PC2
Yield

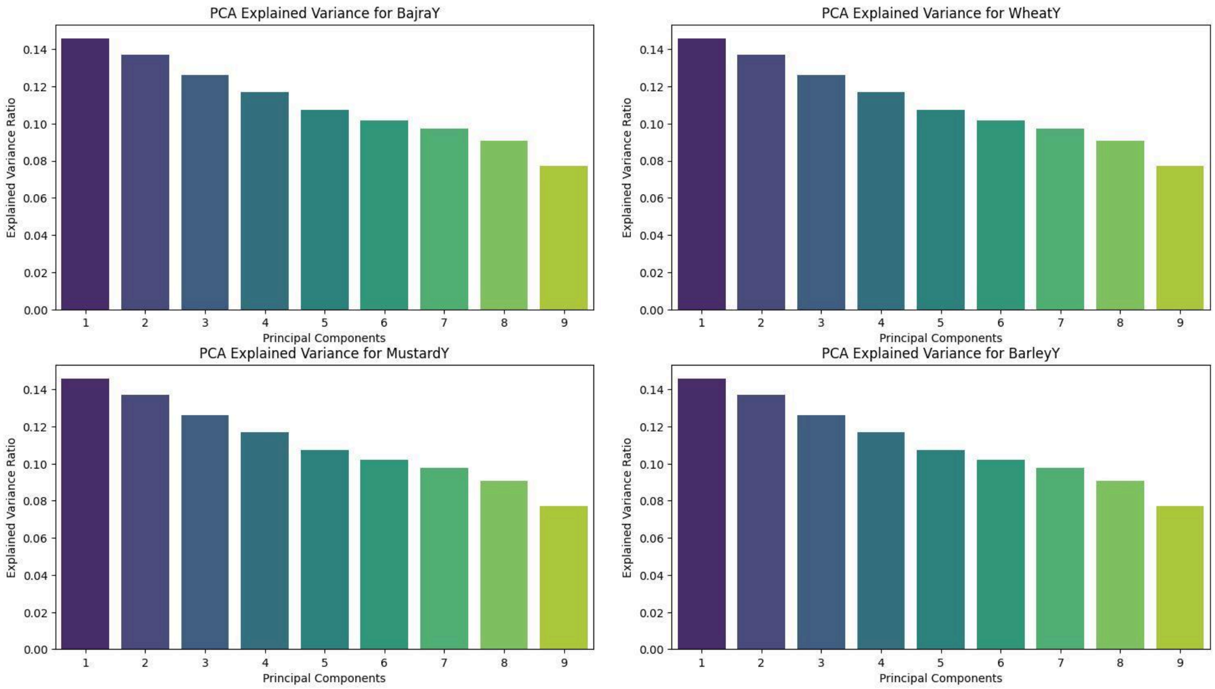

PCA Scatter Plot for BarleyY
Cu
Mn
Zn
Fe
N
P
PC1
PC2
Yield


PCA Scatter Plot for BajraY
PC1
PC2
Yield


PCA Scatter Plot for WheatY
PC1
PC2
Yield


PCA Scatter Plot for MustardY
PC1
PC2
Yield


PCA Scatter Plot for BarleyY
PC1
PC2
Yield



IDEAS CONDENSED AND INTEGRATED

Macro-micro Nutrients Based Soil Fertility Zone Mapping and Assessment Using Fuzzy Logic and Geospatial Analysis (*done by Meeniga*)

- - what fuzzy logic?
- - what geospatial analysis?
- - metric for soil fertility?
- - how does zone mapping help us? (if I plough again tomorrow, zone mapping is of no use then)

AI framework to analyze soil fertility (reference to compare?)

- - define ai framework
- - define soil fertility (what is the metric for this?)
- - what does analysis involve? just geospatial distribution?

AI-based farm-specific crop recommender system for different agroclimatic conditions to minimize crop loss and maximize profit (recommendation based on what?)

- - *define agroclimatic conditions*
  classify datapoints based on quality and quantity of data available online and see if

prediction models are good or not
can we create groups of agroclimatic conditions and use them instead of being very specific to a region? **[NOPE, READ FIRST THREE WORDS]**

- - *objectives: minimizing 'crop loss' and 'maximizing profit*
  define crop loss
  define profit (with respect to what? market prices?)
- - *what does recommendation involve?*
  number of crops to be planted?
  what crops to be planted? (REFER TO P1 and P2)
  quantity of each crop to be planted?
  extra input to be provided for other crops to be taken into consideration while calculating profit

Soil fertility & soil quality evaluation and identifying constraints of micronutrients in optimizing agricultural production

- - define soil fertility
- - define reference for soil quality evaluation
- - like finding limiting factors in a chemical process (becomes crop specific and farm specific).

Ideas for soil quality evaluation:

APPROACH ONE:

- value within suitable range = 1,
- value below range = -1,
- value above given range = 0
- get score for soil fertility, if negative, bad quality, if positive, good

[cross check scoring for within and above range values for each nutrient based on how they affect the plant and is there a detrimental effect in high concentrations] (REFER TO PERSONAL NOTES P4)

APPROACH TWO

- Order micronutrients based on priority of crop type
- Derive a weighted sum for the normalised values (MCDM)
- Evaluate scores.

PERSONAL NOTES

P1: We need to be specific with which crops we are choosing, because we could look at synergistic effects of growing crops together maybe.

P2: look at impact of growing different crops together (competition for resources)

P3: how are we creating zones for different crops? (is it different every time seeds are sown? Becomes a huge hassle for the farmer then)

again, are zones very clearly specified in the farm beforehand and we just choose which ones to grow for which crop (introduces constraints in the equation to be linearly optimized)

P4: benchmarks for quality are set based on impact of the variable (here on health of the plant), and can find suitable ranges for each crop based on some research paper.

check impact at various levels
check mean values in that particular type of soil
check for mean and variance of the values in the target soil

REJECTED IDEAS (to not 'wow' if came across again)

A Cloud-based database of spatial and non-spatial data (just ground survey and upload information, meh)

EASY TO FIND DATA FOR (api based/get database)

temperature, rainfall, humidity, wind, solar radiation
(try using visual crossing api; Openweather needs a subscription, freemium has very less data available)

topographical features like altitude, slope and aspect can be derived from google maps (use osmnx and folium libraries)

growing season is very crop specific, take historical data of what is usually grown in the region (there will some national database for this shit)

access to surface water is from google maps

HARD TO FIND DATA FOR (needs ground survey/ ill maintained database)

check for underground water resources bank (again some random database/ maybe on ground survey)

soil fertility has to be determined through on ground survey, no exception to that. We can only have a generalised idea of what kind of soil is prevalent in that region, not the fertility. At max a rough estimation (impractical big range to just show) of the fertility ranges based on some historical data.

Look into any regular (annual or biannual crop disease or pest infestation critical to the area). Again this becomes very region specific; best derived from on ground survey. or, sigh, historical records, if any.

RELEVANT THEORY

Agroclimatic conditions refer to the climatic factors that directly affect agricultural practices and crop growth in a particular region. These conditions help classify regions based on their climate, soil, and topography, which in turn guides farmers on what crops to grow and when.

Climate Factors:

- Temperature: Influences seed germination, plant growth, and crop yield.
- Rainfall (Precipitation): Determines water availability for crops and irrigation needs.
- Humidity: Affects plant transpiration, disease prevalence, and crop health.
- Wind: Impacts pollination, seed dispersal, and erosion control.
- Solar Radiation: Drives photosynthesis and plant development

Topography:

- Altitude: Higher altitudes often have cooler temperatures and different crop suitability.
- Slope: Impacts water drainage and erosion risk.
- Aspect: The direction a slope faces affects sunlight exposure and microclimate.

Growing Season: The length of time suitable for crop growth, determined by frost dates, temperature ranges, and rainfall patterns.

Water Availability: Access to surface water (rivers, lakes) or groundwater for irrigation.

Soil Characteristics:

- Soil type: Sandy, loamy, clayey — influences water retention and nutrient availability.
- Soil fertility: Determines crop yield potential based on nutrient content (N, P, K levels).
- pH levels: Affects nutrient absorption by plants.

Biotic Factors: Pest and disease prevalence, which is often influenced by the climate and affects crop choice and management practices.

Existing Soil Health Indices

1. USDA Soil Quality Index (SQI)
    a. Developed by the USDA Natural Resources Conservation Service (NRCS).
    b. Evaluates soil health based on a combination of chemical (N, P, K, pH, organic matter), physical (texture, compaction), and biological (microbial activity) indicators.
    c. Uses scoring functions to assign values to each parameter and create an overall score.

📌 Reference: Doran & Parkin (1994) - Soil Health Indicators Framework.

2. Soil Fertility Index (SFI)
    a. Focuses on the availability of essential nutrients (N, P, K, S, Zn, Cu, Fe, Mn, B).
       Uses a weighted scoring system based on nutrient sufficiency levels.
       Typically categorized as:
       High (>80%) – Excellent fertility
       Medium (50-80%) – Moderately fertile
       Low (<50%) – Deficient and needs amendments

📌 Example: Used in FAO and Indian Soil Health Card Program.

3. Comprehensive Soil Health Assessment (Cornell Soil Health Test - CSHS)
    a. Developed by Cornell University.
    b. Evaluates soil health across biological, physical, and chemical properties.
    c. Scoring system from 0 (poor) to 100 (excellent) based on key indicators.

📌 Reference: Moebius-Clune et al. (2016) - Cornell Soil Health Manual.

4. Soil Management Assessment Framework (SMAF)
    a. A quantitative approach developed by the USDA.
    b. Uses a scoring algorithm to convert soil test values into relative quality scores.
    c. Consider chemical properties like nutrient content (NPK, pH, organic matter).

📌 Reference: Karlen et al. (2008) - Applied Soil Ecology.

5. Indian Soil Health Card Scheme
    a. Used in India's National Soil Health Program.
    b. Measures 12 parameters, including macro and micronutrient content.
    c. Provides fertilizer recommendations based on soil test reports.

📌 Website: https://soilhealth.dac.gov.in/

How You Can Use These Scales
If you want a simple nutrient-based score, the Soil Fertility Index (SFI) is the best fit.
If you need a broader soil health perspective, the USDA SQI, SMAF, or Cornell SHS would be better.

Rabi Crops: Wheat, Mustard, Barley
Kharif Crop: Bajra

For Kharif Crops, source: https://pib.gov.in/PressReleaseIframePage.aspx?PRID=2026698
For Rabi Crops, source: https://pib.gov.in/PressReleasePage.aspx?PRID=2065310

Crop	MSP (2025-26)	Cost of Production (2025-26)
Wheat	24.25	11.82
Barley	19.80	12.39

Mustard	59.50	30.11
Bajra	26.25	14.85

NOTE: Values for Bajra are of 2024-25

Cost includes all paid out costs such as those incurred on account of hired human labour, bullock labour/machine labour, rent paid for leased in land, expenses incurred on use of material inputs like seeds, fertilizers, manures, irrigation charges, depreciation on implements and farm buildings, interest on working capital, diesel/electricity for operation of pump sets etc., miscellaneous expenses and imputed value of family labour.

ABSTRACT

Crop diversification emerges as a transformative strategy in sustainable agriculture, offering a pathway to ecological stability, enhanced farm productivity, and economic security for farmers. The scientific challenge lies not only in recommending alternative crops but in creating a data-driven framework that harmonizes soil characteristics, fertility levels, and economic variables to produce precise, profit-oriented strategies.

Our study pioneers an AI-powered approach through KHETI (Knowledge-driven Harvest for Ecological Transformation and Innovation), a model designed to assess soil quality and fertility and propose crop combinations that maximize profit potential and mitigate crop failure risks. Unlike existing methodologies, KHETI dynamically bridges scientific soil insights with real-time market data, offering hyper-local, customized farming strategies.

The model draws from comprehensive ground survey reports, incorporating micronutrient profiles, sample coordinates, historical crop yields, and prevailing market prices for key crops like wheat, barley, bajra, and maize. It further accounts for critical parameters such as water availability, underground water levels, and regional pest and disease patterns. At its core, the study employs optimization in the profit equation subject to the constraints of designated farm area, predicted yield, cost of production and market prices..

This research has transformative implications—empowering farmers with data-centric crop planning, guiding policymakers to endorse sustainable practices, and equipping scientists with quantifiable insights for ecological advancement.