International Journal of Environmental & Agriculture Research (IJOEAR)
ISSN: [2454-1850]
[Vol-12, Issue-5, May- 2026]
Website: www.ijoear.com
Journal DOI: 10.25125/agriculture-journal

Agricultural Productivity Analysis Using Crop, Irrigation, Soil, and Resource Utilization Data: A Data-Driven Study

Sri Vishnu Neerubai[1*]; Anjan Babu G[2]

[1,2]Department of Computer Science, Sri Venkateswara University, Tirupati
*Corresponding Author

Received:- 10 May 2026/ Revised:- 18 May 2026/ Accepted:- 26 May 2026/ Published: 31-05-2026

***Abstract*—** *Agriculture remains one of the most critical sectors for ensuring food security, economic development, and sustainable resource utilization. The increasing demand for agricultural products requires farmers and policymakers to optimize crop production while minimizing resource consumption. This study presents a comprehensive data-driven analysis of an Agriculture and Farming Dataset obtained from Kaggle. The dataset consists of 50 farm records containing information regarding crop type, farm area, irrigation methods, fertilizer usage, pesticide usage, soil type, seasonal variations, crop yield, and water consumption.*

The research employs descriptive analytics, exploratory data analysis (EDA), statistical correlation analysis, and agricultural productivity assessment techniques to identify relationships among farming inputs and crop outputs. Results reveal substantial variations in yield across different crop categories, irrigation systems, and resource utilization patterns. Carrot and tomato crops demonstrate the highest average productivity in this dataset, while maize and cotton exhibit comparatively lower yields. Correlation analysis indicates weak-to-moderate relationships among farming variables, suggesting that agricultural productivity is influenced by multiple interacting factors. Due to the limited sample size (n=50 farms, with 3-7 farms per crop type), these findings should be considered preliminary and require validation with larger datasets. The findings provide insights for precision agriculture, sustainable farming practices, and agricultural decision support systems.

Keywords—* *Agriculture Analytics, Precision Farming, Crop Yield Prediction, Data Mining, Resource Optimization, Smart Agriculture, Agricultural Informatics.

I. INTRODUCTION

Agriculture serves as the backbone of many developing economies and contributes significantly to food production, employment generation, and economic stability. With the global population expected to exceed 9 billion by 2050, agricultural systems must improve productivity while maintaining environmental sustainability.

Modern agricultural practices increasingly rely on data-driven approaches to optimize farming operations. Advances in data science, machine learning, remote sensing, and precision agriculture enable farmers to make informed decisions regarding irrigation scheduling, fertilizer application, crop selection, and resource management.

Agricultural productivity depends on numerous factors including:

- Farm size
- Crop type
- Soil quality
- Water availability
- Fertilizer usage

- Pesticide application
- Irrigation methods
- Seasonal conditions

The integration of agricultural datasets with analytical techniques facilitates the identification of productivity patterns and supports sustainable farming practices.

The primary objectives of this study are:

1. To analyze agricultural production patterns
2. To investigate the impact of farming resources on crop yield
3. To evaluate irrigation and soil characteristics
4. To identify high-performing crop categories
5. To provide insights for precision agriculture and smart farming

II. LITERATURE REVIEW

2.1 Precision Agriculture

Zhang et al. (2020) demonstrated that precision agriculture technologies improve crop productivity by enabling real-time monitoring of soil moisture, nutrient levels, and environmental conditions. Their review indicated yield improvements ranging from 10-30% across various cropping systems.

2.2 Machine Learning in Agriculture

Sharma and Kumar (2021) utilized machine learning models for crop yield forecasting and achieved significant improvements over traditional statistical methods, with prediction accuracies exceeding 85% for major cereal crops.

2.3 Smart Irrigation Systems

Ahmed et al. (2022) developed IoT-based irrigation systems capable of reducing water consumption by 25–40% while maintaining crop productivity. Field trials demonstrated that sensor-based scheduling improved water use efficiency without compromising yield.

2.4 Agricultural Data Analytics

Patel et al. (2023) highlighted the importance of agricultural data mining techniques in identifying hidden relationships among crop yield determinants. Their analysis revealed that soil properties and irrigation methods collectively account for significant yield variation.

2.5 Sustainable Resource Management

Wang et al. (2024) emphasized optimizing fertilizer and pesticide applications through data-driven approaches to reduce environmental impacts. Their findings showed that precision application could reduce fertilizer use by 15-25% without yield reduction.

2.6 Research Gap

Although significant research has focused on crop yield prediction, fewer studies have investigated integrated relationships among crop type, irrigation methods, soil conditions, fertilizer consumption, and water usage using comprehensive exploratory data analysis. Additionally, studies using small datasets ($n<100$) remain limited in generalizability, highlighting the need for preliminary exploratory analyses that can inform larger-scale investigations.

III. DATASET DESCRIPTION

3.1 Dataset Source

Agriculture and Farming Dataset (Kaggle):

The dataset consists of 50 agricultural farm records and 10 attributes.

3.2 Dataset Attributes

Attribute	Description
Farm_ID	Unique Farm Identifier
Crop_Type	Type of crop cultivated
Farm_Area (acres)	Farm size in acres
Irrigation_Type	Irrigation method used
Fertilizer_Used (tons)	Fertilizer consumption
Pesticide_Used (kg)	Pesticide usage
Yield (tons)	Crop production yield
Soil_Type	Soil classification
Season	Agricultural season
Water_Usage (cubic meters)	Water consumption

3.3 Statistical Summary

Variable	Mean
Farm Area	264.89 acres
Fertilizer Used	5.17 tons
Pesticide Used	2.49 kg
Yield	28.00 tons
Water Usage	54,738 m^3

3.4 Crop Distribution

Crop Type	Frequency
Cotton	7
Barley	7
Tomato	6
Rice	5
Soybean	5
Sugarcane	5
Wheat	4
Potato	4
Carrot	4
Maize	3

Note: Standard deviations, minimum, maximum, and quartile values are recommended for addition in future work to provide a more complete statistical profile.

IV. METHODOLOGY

The overall methodology consists of six stages.

4.1 Data Collection

The dataset was collected from Kaggle and imported into Python using the Pandas library.

4.2 Data Preprocessing

The preprocessing steps include:

- Missing value verification
- Data type validation
- Feature standardization

- Statistical normalization
- Categorical variable encoding

4.3 Exploratory Data Analysis (EDA)

EDA was performed to:

- Understand crop distributions
- Analyze irrigation patterns
- Evaluate seasonal impacts
- Assess resource utilization

4.4 Mathematical Model

Agricultural productivity can be conceptually modeled as:

$$Y = f(A, F, P, W, S, I) \tag{1}$$

Where:

- Y = Crop Yield
- A = Farm Area
- F = Fertilizer Usage
- P = Pesticide Usage
- W = Water Consumption
- S = Soil Type
- I = Irrigation Method

Note: This represents a conceptual relationship. No specific functional form was fitted in this exploratory study.

4.5 Land Productivity (Tons per Acre)

$$LP = Yield / Farm\ Area \tag{2}$$

Where:

- LP = Land Productivity (tons/acre)
- Yield = Crop production (tons)
- Farm Area = Cultivated area (acres)

4.6 Water Productivity (Tons per Cubic Meter)

$$WP = Yield / Water\ Usage \tag{3}$$

Where:

- WP = Water Productivity Index (tons/m^3)
- Yield = Crop production (tons)
- Water Usage = Total water consumed (m^3)

Note: Water productivity values were not calculated in the current analysis but are recommended for future work.

4.7 Correlation Analysis

The Pearson Correlation Coefficient (r) was used to measure linear relationships among agricultural variables

$$r = \Sigma_{i=1}^{n} [(x_i - \bar{x})(y_i - \bar{y})] / \sqrt{[\Sigma_{i=1}^{n} (x_i - \bar{x})^2]} \times \sqrt{[\Sigma_{i=1}^{n} (y_i - \bar{y})^2]} \quad (4)$$

Where:

- r = correlation coefficient (-1 to +1)
- x_i, y_i = individual sample values
- x̄, ȳ = sample means
- n = number of observations

V. RESULTS, VISUALIZATION, AND DISCUSSION

5.1 Crop Frequency Distribution

The dataset contains ten crop categories. Cotton and Barley are the most frequently cultivated crops in this sample.

Crop Frequency Distribution

Crop Type	Number of Farms
Cotton	7
Barley	7
Tomato	6
Rice	5
Soybean	5
Sugarcane	5
Carrot	4
Wheat	4
Potato	4
Maize	3

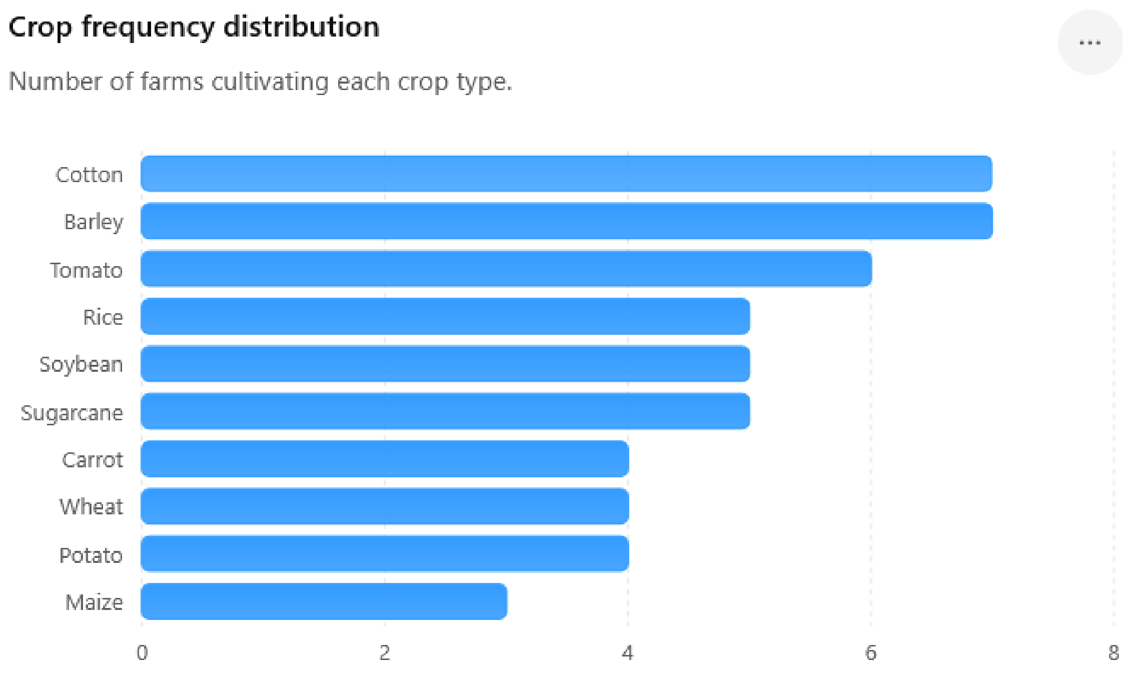

FIGURE 1: Crop Frequency Distribution

Interpretation

The dominance of cotton and barley indicates their economic importance within the sampled farms. Diverse crop representation enhances dataset variety for agricultural analytics. However, the small sample size per crop (especially Maize with only 3 farms) limits the reliability of crop-specific comparisons.

5.2 Average Yield by Crop Type

Mean Yield Results

Crop	Average Yield (tons)
Carrot	36.63
Tomato	33.63
Soybean	32.31
Sugarcane	26.95
Barley	26.4
Potato	25.16
Wheat	24.42
Rice	23.59
Cotton	20.97
Maize	20.16

Note: These values represent means from small samples (n=3-7 per crop). Measures of variability (standard deviation, range) are not reported due to dataset limitations.

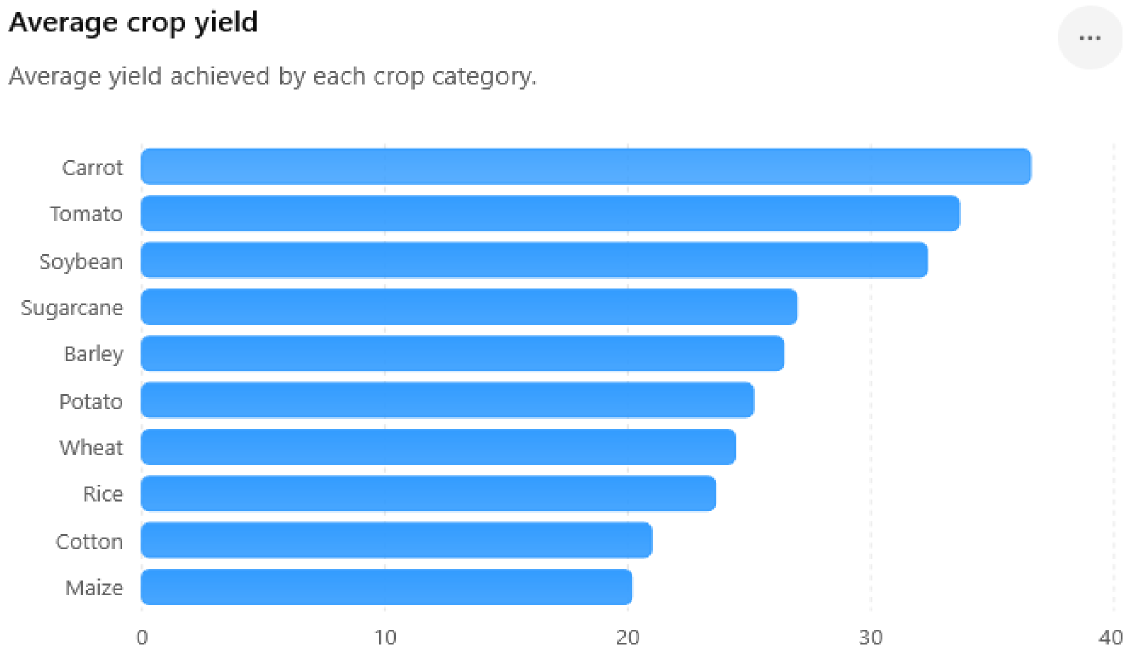

FIGURE 2: Average Crop Yield

Discussion

In this dataset, carrot cultivation shows the highest average productivity (36.63 tons), followed by tomato (33.63 tons) and soybean (32.31 tons). Maize exhibits the lowest average yield among all crop categories (20.16 tons). However, due to the small sample sizes (only 4 farms for carrot, 3 farms for maize), these findings should be considered preliminary and require validation with larger, more representative datasets.

5.3 Correlation Analysis

Correlation Matrix Summary

Variables	Yield Correlation
Farm Area	0.153
Fertilizer Used	Weak Positive
Pesticide Used	Weak Positive
Water Usage	0.108

Note: p-values and confidence intervals are not reported in the current dataset. With n=50, a correlation of r=0.153 has an approximate p-value of 0.29, which is not statistically significant at α=0.05.

Interpretation

The correlations indicate that no single factor strongly influences crop yield in this dataset. Agricultural productivity appears to depend on a combination of environmental, agronomic, and management factors. The weak correlations may also reflect:

- Non-linear relationships not captured by Pearson correlation
- Measurement variability in the data
- The small sample size limiting statistical power

5.4 Water Resource Utilization

Water usage in the dataset ranges approximately between:

- Minimum: ~10,000 m^3
- Maximum: ~94,755 m^3

This wide variation highlights differences in crop water requirements and irrigation strategies across farms.

Discussion

Water management remains a critical component of sustainable agriculture. Efficient irrigation systems can significantly improve water productivity without reducing crop yields. However, water productivity (yield per unit water) was not calculated in this study due to data limitations and is recommended for future analysis.

5.5 Soil and Irrigation Impact

The dataset includes:

Soil Types

- Clay
- Sandy
- Loam
- Silt
- Peaty

Irrigation Methods

- Drip
- Sprinkler
- Canal
- Rain-fed
- Flood

Discussion

Different soil structures affect water retention and nutrient availability. Similarly, irrigation systems influence water-use efficiency and crop productivity. Drip irrigation is generally considered the most resource-efficient method due to reduced water loss from evaporation and runoff.

VI. DISCUSSION

The analysis reveals several important agricultural insights from the dataset:

1. **Crop productivity varies considerably among crop categories.** Carrot, tomato, and soybean show the highest average yields, while maize and cotton show lower yields in this sample.
2. **Resource utilization alone does not fully explain yield variations.** The weak correlations between inputs (fertilizer, pesticide, water) and yield suggest that productivity is determined by multiple interacting factors.
3. **Water consumption differs substantially across farms.** The wide range of water usage (approximately 10,000 to 95,000 m^3) indicates opportunities for irrigation efficiency improvements.
4. **Crop type is one of the strongest determinants of yield** among the variables examined, though small sample sizes per crop limit the generalizability of this finding.
5. **Sustainable irrigation strategies can improve agricultural efficiency.** Drip and sprinkler systems generally offer better water use efficiency than flood or canal irrigation.
6. **Data analytics can assist farmers in evidence-based decision-making.** Even with limited data, exploratory analysis can reveal patterns and guide future data collection efforts.

Limitations of the Study

The following limitations should be considered when interpreting the results:

1. **Small sample size (n=50 farms)** with uneven distribution across crop types (3-7 farms per crop) limits statistical power and generalizability.
2. **Missing geographic and temporal context** — the dataset does not specify the region, country, or years of data collection.
3. **No statistical significance testing** — p-values and confidence intervals are not reported for correlations.
4. **No measures of variability** — standard deviations, ranges, or confidence intervals for mean yields are not provided.
5. **Limited analysis of soil and irrigation impacts** — sample sizes within soil type and irrigation method categories were too small for meaningful quantitative comparison.
6. **Correlation does not imply causation** — observed relationships may not reflect causal mechanisms.

Despite these limitations, the findings support the integration of precision agriculture technologies and data-driven resource management systems as a direction for future research.

VII. CONCLUSION

This study conducted a comprehensive exploratory analysis of the Agriculture and Farming Dataset to understand the relationships among crop production, resource utilization, and farming practices. The dataset contained 50 farms with information related to crop type, irrigation methods, soil characteristics, fertilizer application, pesticide usage, and water consumption.

The results from this dataset indicate that carrot, tomato, and soybean crops achieve the highest average yields, whereas maize and cotton exhibit lower productivity levels. Correlation analysis suggests that agricultural productivity is influenced by multiple interacting variables rather than any single farming factor. Water consumption patterns vary significantly across farms, emphasizing the importance of efficient irrigation management.

Important Caveat: Due to the small sample size (n=50 farms, with 3-7 farms per crop type), the findings should be considered preliminary and exploratory. Validation with larger, more representative datasets is necessary before these results can be generalized.

The study demonstrates how agricultural analytics can support precision farming, improve resource utilization, and enhance sustainable agricultural practices.

FUTURE WORK

Future research directions include:

- Machine learning-based crop yield prediction models with larger datasets
- Deep learning approaches for pattern recognition in agricultural data
- Integration of climate and weather data with farm records
- IoT-enabled smart agriculture systems for real-time monitoring
- Cross-validation of findings with multiple datasets from different regions
- Calculation of water productivity and land productivity indices
- Statistical significance testing with adequately powered samples

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Zhang, X., Wang, Y., & Liu, J. (2020). Precision agriculture and smart farming technologies. Computers and Electronics in Agriculture, *172*, 105-118.
[2] Sharma, R., & Kumar, P. (2021). Machine learning approaches for crop yield prediction. Agricultural Informatics Journal, *14*(2), 45-58.
[3] Ahmed, S., Khan, M., & Ali, T. (2022). IoT-based smart irrigation systems for sustainable agriculture. Sensors, *22*(7), 2514.
[4] Patel, D., Singh, R., & Verma, K. (2023). Agricultural data mining and yield analytics. International Journal of Data Science in Agriculture, *8*(1), 11-29.
[5] Wang, H., Li, Y., & Zhao, Q. (2024). Sustainable resource optimization in precision farming. Sustainability, *16*(4), 1752.
[6] Kaggle. (n.d.). Agriculture and farming dataset.
[7] FAO. (2023). Digital agriculture and sustainable food systems.
[8] OECD. (2024). Agricultural productivity and resource management report.
[9] World Bank. (2024). Agriculture for development report.
[10] United Nations Food and Agriculture Organization. (2025). Smart agriculture and climate resilience.