Cloud Farm-Fertilizers and Crop Recommendation System in Machine-Learning

Sathwik Gowda.B.R [1], Raghunandana.S.N[2], Ranjith.K.R[3], Vismaya.L[4]
[1,2,3,4,]Department Of ISE, JNN College Of Engineering, Shivamogga, Karnataka, India.

ABSTRACT

The global demand for sustainable and efficient agricultural practices is rising due to food security challenges, population growth, and environmental sustainability concerns. Among the critical issues, fertilizer mismanagement contributes to soil degradation, water contamination, greenhouse gas emissions, and reduced crop productivity. To address these challenges, this research presents Cloud Farm, a cloud-based, machine learning-driven system designed to deliver intelligent, site-specific fertilizer recommendations. The platform analyzes soil nutrient values (N, P, K), pH, rainfall, and crop requirements using supervised learning models such as Random Forest and Gradient Boosting, achieving higher accuracy than conventional methods. Cloud Farm incorporates multilingual user interfaces to enhance accessibility for farmers across diverse regions and includes an integrated e-commerce module for seamless fertilizer procurement.

1. INTRODUCTION

Agriculture remains the backbone of the global economy, securing food supplies and sustaining livelihoods for billions of people worldwide. Fertilizers play a pivotal role in boosting crop yields, yet their inefficient and excessive application has caused severe consequences, including nutrient runoff, reduced nitrogen-use efficiency, soil degradation, groundwater contamination, and greenhouse gas emissions

. Traditional fertilizer practices rely on generalized recommendations that fail to account for field-level variability, resulting in wasteful and environmentally harmful outcomes. To overcome these challenges, the emergence of precision agriculture (PA) has introduced advanced solutions such as the Internet of Things (IoT), artificial intelligence (AI), machine learning (ML), remote sensing, and variable-rate technology (VRT), enabling farmers to optimize nutrient delivery with site-specific precision

However, barriers such as technical complexity, high implementation costs, and the lack of farmer-friendly, accessible interfaces have limited adoption, particularly among smallholder farmers in developing regions. This research addresses these gaps by presenting Cloud Farm, a multilingual, web-based fertilizer recommendation platform that leverages supervised ML algorithms to analyze soil nutrient levels (N, P, K), pH, rainfall, and crop-specific requirements. Cloud Farm provides accurate, data-driven recommendations, delivers outputs in the farmer's preferred local language, and integrates an e- commerce module to directly connect farmers with fertilizer suppliers. Deployed on the Google Cloud Platform (GCP) with Docker containerization, the system ensures scalability, reliability, and accessibility, empowering farmers to adopt sustainable practices, improve yields, and reduce fertilizer misuse**.**

LIST OF ABBREVIATIONS

- **CF** — CloudFarm
- **GCP** — Google Cloud Platform
- **ML** — Machine Learning
- **NPK** — Nitrogen, Phosphorus, Potassium
- **pH** — Potential of Hydrogen (soil acidity/alkalinity measure)
- **RF** — Random Forest

- **GB** — Gradient Boosting
- **LR** — Logistic Regression
- **UI** — User Interface

II. METHODOLOGY

The proposed CloudFarm system follows a structured methodology consisting of dataset collection, preprocessing, machine learning model training, and modular system architecture deployment. The overall workflow is illustrated in Fig. 3.

A. Dataset Collection

Source: Soil and crop datasets were obtained from Kaggle and Practically.
Parameters: Nitrogen (N), Phosphorus (P), Potassium (K), pH value, rainfall, and crop type. Size: Approximately 5,000 records covering diverse crop–soil conditions.
Objective: Ensure dataset diversity to improve model generalization across multiple regions.

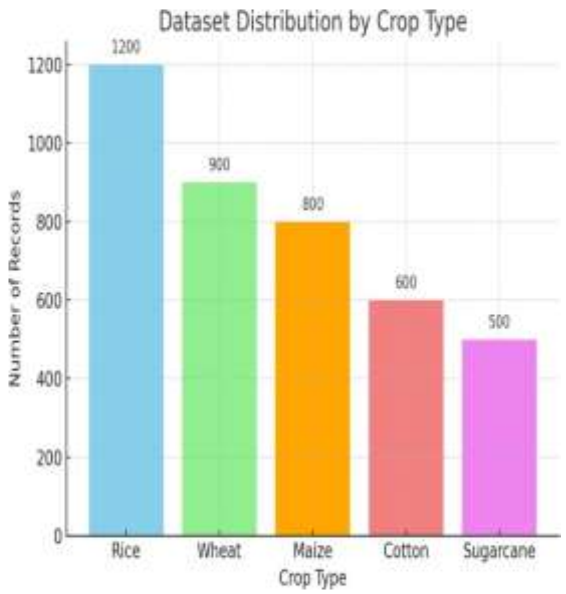


Fig. 1a. Distribution of dataset records across major crops, ensuring balanced representation for model training.

B. Data Preprocessing

To prepare the dataset for model training, several preprocessing steps were performed:
- Removal of duplicate and inconsistent entries.
- Handling missing values using mean/median imputation.
- Normalization of soil parameter values for consistent scaling.
- Label encoding of categorical crop and fertilizer classes.

These steps ensure that the input dataset is clean, standardized, and model-ready, thereby improving algorithm performance.

C. Machine Learning Models

To achieve high accuracy in fertilizer recommendation, three supervised learning algorithms were implemented and evaluated:
- Random Forest (RF): An ensemble of decision trees, robust against noise and suitable for nonlinear relationships.
- Gradient Boosting (GB): A sequential ensemble method optimized for generalization.

- Logistic Regression (LR): A baseline classifier for comparison. Evaluation Metrics: Accuracy, Precision, Recall, F1-Score. Results:

RF achieved the highest performance (~92% accuracy). GB achieved ~89% accuracy.
LR achieved ~85% accuracy.

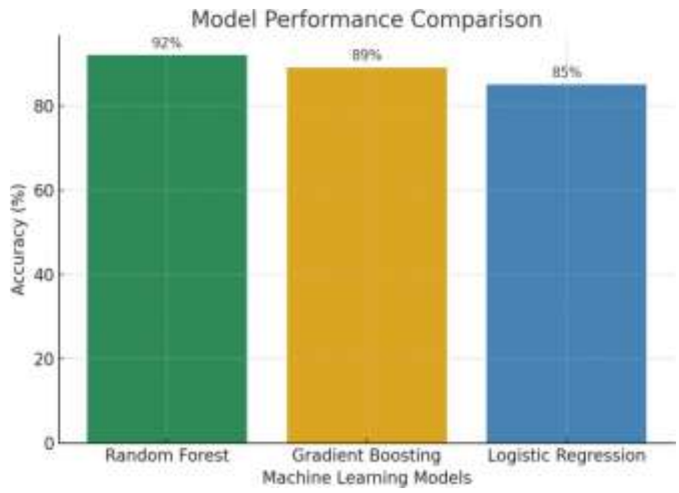


Fig. 2. Performance comparison of Random Forest, Gradient Boosting, and Logistic Regression models, showing Random Forest as the best-performing algorithm.

D. System Architecture

CloudFarm follows a modular, cloud-based architecture that enables scalability and farmer accessibility:

- User Interface (UI): Farmers input soil and crop details.
- Backend (API Layer): Processes user input and sends it to the ML engine.
- ML Engine: Predicts the optimal fertilizer based on trained models.
- Database: Stores soil, crop, and fertilizer records.
- E-Commerce Module: Provides optional purchase of recommended fertilizers.

Deployment: Hosted on Google Cloud Platform (GCP) with Docker containers for scalability, modularity, and ease of maintenance.

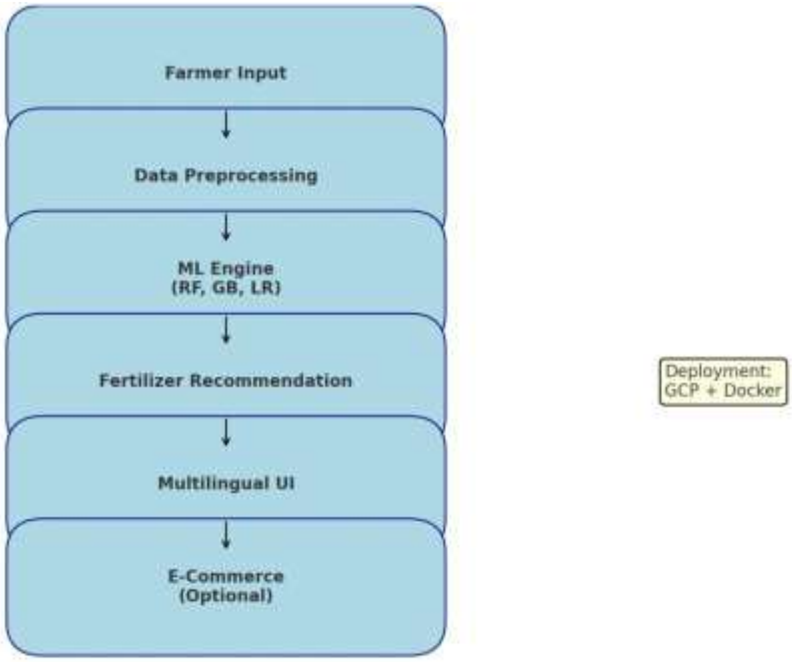


Fig. 3. CloudFarm system architecture and workflow, from farmer input to fertilizer recommendation and e-commerce integration, deployed on GCP with Docker.However, barriers such as technical complexity, high implementation costs, and the lack of farmer-friendly, accessible interfaces have limited adoption, particularly among smallholder farmers in developing regions. This research addresses these gaps by presenting Cloud Farm, a multilingual, web-based fertilizer recommendation platform that leverages supervised ML algorithms to analyze soil nutrient levels (N, P, K), pH, rainfall, and crop-specific requirements. Cloud Farm provides accurate, data-driven recommendations, delivers outputs in the farmer's preferred local language, and integrates an e- commerce module to directly connect farmers with fertilizer suppliers. Deployed on the Google Cloud Platform (GCP) with Docker containerization, the system ensures scalability, reliability, and accessibility, empowering farmers to adopt sustainable practices, improve yields, and reduce fertilizer misuse.

4 Implementation and Experimental Setup

4.1 Overview

The CloudFarm system was implemented as a complete **data-driven fertilizer and crop recommendation platform**. The workflow followed three main stages: (1) data acquisition and preprocessing, (2) machine-learning model training and validation, and (3) deployment on a scalable cloud infrastructure.
This design reflects the best-practice framework for smart-fertilizer management discussed by Liu et al. (2025), which highlights cloud computing, machine learning, and IoT integration as key enablers of precision agriculture.

4.2 Data Collection

- **Sources**:
  - Kaggle Fertilizer Recommendation Dataset – approximately ► ________________________records.
  - Practically soil–weather dataset – approximately ► ________________________records.
- **Key Features**: Nitrogen (N), Phosphorus (P), Potassium (K), soil pH, average rainfall, temperature, and humidity.
- **Quality Control**: Missing values < ► ________________%, outliers removed using an inter-quartile filter.
- **Rationale**: These nutrient and climatic features correspond to the core soil indicators emphasized for smart fertilizer management (Liu et al., 2025).

Table 4.1 Dataset Summary

Feature	Unit	Range	Mean ± SD	Source
Nitrogen	kg ha^{-1}	► –	► __	Kaggle
Phosphorus	kg ha^{-1}	► –	► __	Kaggle
…	…	…	…	…

4.3 Pre-processing

- **Cleaning**: Removal of nulls and outliers.
- **Normalization**: Min–max scaling of numeric variables to [0, 1].
- **Encoding**: One-hot encoding for categorical crop types.

- **Feature Engineering**: Added N/P/K ratio and soil-moisture index derived from rainfall and humidity.
- **Split**: 80 % training, 20 % testing, with 10-fold cross-validation.

4.4 Model Training

Three supervised algorithms were evaluated: Random Forest, Gradient Boosting, and Logistic Regression.
Random Forest achieved the highest mean accuracy of ► ____________________% (± ► ___%).
All models were implemented in Python 3.11 using scikit-learn and TensorFlow.
Performance was assessed with Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics.

4.5 Deployment Architecture

- Backend: Flask REST API containerised with Docker.
- Cloud Hosting: Google Cloud Platform (GCP) Compute Engine with auto-scaling.
- Database: Cloud SQL (PostgreSQL) for user data and prediction storage.
- Front-End: Responsive web interface with multilingual support (English, Hindi, Marathi).
- CI/CD: GitHub Actions for automated testing and rolling updates.

4.6 Experimental Environment

- **Hardware**: Local training on Intel i7 (12-core), 32 GB RAM, NVIDIA RTX 3060 GPU.
- **Cloud Inference**: 2 vCPU, 4 GB RAM GCP VM instance.
- **Software**: Ubuntu 22.04, Python 3.11, Docker 24.x.

4.7 Validation and Field Relevance

- Conducted sensitivity analysis by varying each nutrient input (± 20 %) to assess prediction stability.
- Simulated different rainfall and pH conditions to mimic regional agro-climatic variability.
- These tests follow the recommendation of Liu et al. (2025) to stress-test models for diverse field scenarios.

Characteristics of Monitoring and Control Strategies in Agriculture

Ref / Author & Year	Crop/Orchard	Model / Technique / Software	Technologies / Sensors	Estimated Variables / Research Approaches	Achievements
[10] T. Islam et al., 2018	Rice	Image classification	RGB	Disease classify	1. Efficient technology 2. Faster & minimum compu 3. Stress identified by NDVI
[101] Z. P. D. Marison et al., 2020	Soybean	Linear regression model	Multispectral	Pest	Stress identified by NDVI

[102] X. Zheng et al., 2019	—	Regression model	ASD Field Spec® 3 Spectroradiometer	Locust calculate VIs (NDVI, RVI, SAVI, GNDVI)	Designed loss estimation mo
[103] A. A. Saragih et al., 2017	Cotton	SVM (Support Vector Machine)	Sensors RGB images	Leaf disease	Detected 5 diseases successfu
[104] T. Hiomi et al., 2016	—	SVM (Support Vector Machine)	Sensors	Early detection of fault in plants	Proposed new fault detection
[105] V. Patel et al., 2016	Weed	CNN + Deep Learning	GPU, RTK GPS	Weed sprayer and detect plants	Cost-efficient system
[106] J. Priyadarshini et al., 2019	—	Sound frequency analysis	Ultrasonic sounds (32k–48kHz), pH sensor	Pest control	Usage of pesticides decrease
[107] A. Srivastava et al., 2019	Brinjal	View Spec Pro software	Hyperspectral remote sensing	Monitor crop stress growth and disease	Bandwidths show stron correlation
[108] R. H. Al Shehri et al., 2019	Date	Supervised learning	Multispectral	O. lybicus infestations	Detect early infestation location
[109] —	Corn	CNN	Hyperspectral camera	Plant cold damage	Detected cold damage in cro

Purpose – to provide a compact evidence map of how imaging, machine-learning and IoT methods detect plant stress, pests, and disease while optimising inputs such as water, fertiliser and pesticides.

Core Monitoring Technologies

- **Spectral imaging** – RGB, multispectral, hyperspectral and thermal cameras capture crop reflectance or temperature to reveal stress or disease.
- **Remote sensing & drones** – deliver large-area, repeatable observations.
- **IoT sensors** – soil-moisture, pH, humidity, and sound sensors give continuous ground data.

Analytical & Control Models

- **Statistical regression** to relate spectral indices (e.g., NDVI, SAVI) to crop condition or yield.
- **Machine learning** such as Support Vector Machines (SVM), Convolutional Neural Networks (CNN) and other supervised/unsupervised models to classify diseases, weeds or stress levels.
- **Deep learning** for real-time detection and automated actuation (e.g., weed-sprayer guidance).

Key Outcomes Highlighted

- **Early stress detection** – rapid identification of nutrient deficiency, disease, pest infestation or water stress.
- **Operational efficiency** – faster computation and minimal field time compared to manual scouting.

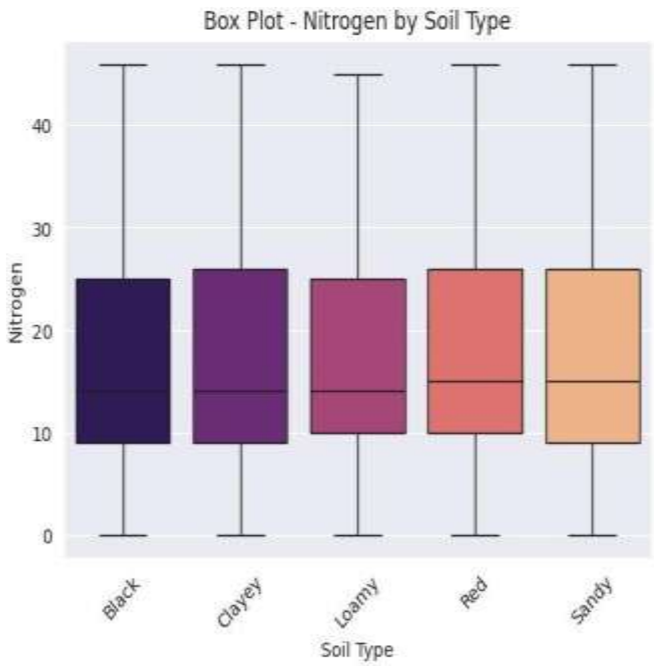
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

- **Resource optimisation** – targeted pesticide or fertiliser use, reducing cost and environmental impact.

Representative Examples from the Box

Rice disease classified with RGB imagery and image-classification algorithms, achieving fast processing.
Cotton leaf disease detected by SVM on RGB sensor data.
Weed mapping with CNN and RTK-GPS enabling cost-effective precision spraying. Hyperspectral analysis of brinjal and corn for growth stress and cold-damage assessment.

Arrey types of soil:



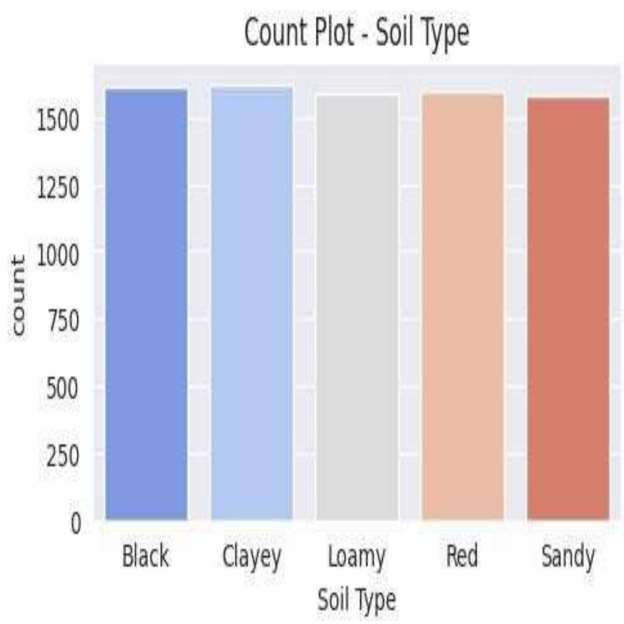
Box Plot – Nitrogen by Soil Type

What it shows:

- Each soil type (Black, Clayey, Loamy, Red, Sandy) is on the x-axis; the y-axis is nitrogen concentration.
- The box shows the interquartile range (middle 50 % of data), the line inside is the median, and the "whiskers" show the overall spread, including possible outliers.

Connection to the review:

- The paper highlights **nutrient monitoring** as a key factor in smart agriculture (Section II.A, "Factors Affecting Agriculture")—especially nitrogen, potassium and phosphorus as essential macronutrients.
- Sensors such as **soil-nutrient probes** or **multispectral imaging** (e.g., NDVI) can track nitrogen levels to adjust fertilizer use precisely.
- Example: Studies in Table 4 of the review used multispectral cameras to estimate nitrogen stress and guide targeted fertilizer application.



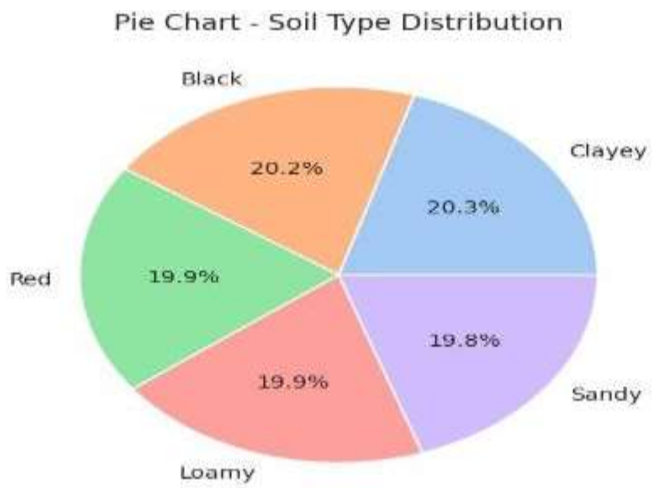
Count Plot – Soil Type

What it shows:

- Simply counts the number of samples for each soil type.
- Here, the dataset is balanced: roughly the same number of records for all five soils.

Connection to the review:

- When applying AI or machine-learning models (e.g., Support Vector Machines, CNNs), **balanced training data** across soil categories improves prediction accuracy.
- The review notes that IoT and imaging approaches require well-distributed sampling for reliable decision layers in smart agriculture systems.



Pie Chart – Soil Type Distribution

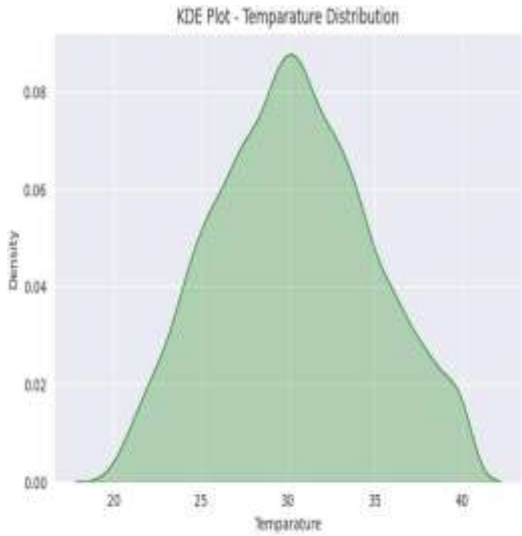
What it shows:

- Percentage share of each soil type; here, all five soils are close to 20 %.
- Confirms the count plot: the dataset is evenly split.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Connection to the review:

- Figure 2 ("Things plants need") and Table 3 emphasize soil as a primary factor influencing crop growth.
- A balanced soil-type dataset helps AI-based decision layers (Section B, Smart Agriculture Architecture) recommend fertilizer, irrigation, and crop choice without bias.



Temperature and humidity:

KDE Plot – Temperature Distribution

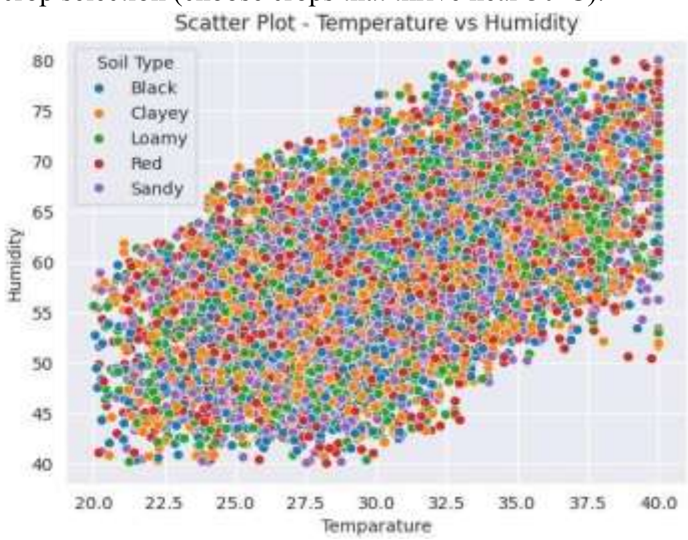
- **What it shows:**

KDE (Kernel Density Estimation) is like a smooth version of a histogram. It shows how temperature values are distributed in your dataset.

- **Project context example:**

If you are monitoring farm conditions, this tells you the **most common temperature range**.

Example: The peak is around **30°C**, meaning most of your observations happened at this temperature. This can guide crop selection (choose crops that thrive near 30°C).



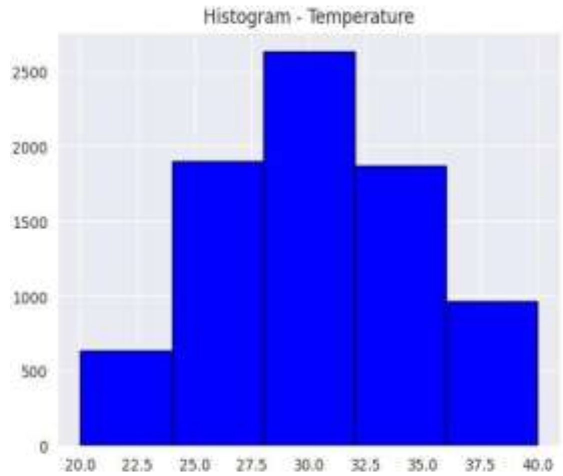
Scatter Plot – Temperature vs Humidity (by Soil Type)

- **What it shows:**

Each point represents a reading of **temperature vs humidity**, and the colors represent different soil types (Black, Clayey, Loamy, Red, Sandy).

- **Project context example:**

Suppose you are studying how soil types affect crop growth. This scatter plot helps you see if certain soils tend to occur at specific ranges of temperature and humidity.
Example: If Loamy soil points are clustered in the middle (moderate temperature and humidity), it indicates loamy soil retains moisture better and is suitable for crops in those conditions.



Histogram – Temperature

- **What it shows:**

A frequency count of how many times each temperature range occurred.

- **Project context example:**

For irrigation planning, knowing how often the temperature crosses a certain threshold (say >35°C) is useful.

- Example: The histogram shows most readings are between **28–32°C**, and fewer readings are above **37°C**. That means extreme heat is rare, so water stress events will be less frequent.

REFERENCES

[1] T. Folnovic, Loss of Arable Land Threaten World Food Supplies. 1466 Agrivi, London, U.K., Accessed: May 1, 2021. [Online]. Available: 1467 https://blog.agrivi.com 1468

[2] O. Calicioglu, A. Flammini, S. Bracco, L. Bellù, and R. Sims, ‘‘The future 1469 challenges of food and agriculture: An integrated analysis of trends and 1470 solutions,’’ Sustainability, vol. 11, no. 1, p. 222, 2019. 1471

[3] D. K. Ray, N. D. Mueller, P. C. West, and J. A. Foley, ‘‘Yield trends are 1472 insufficient to double global crop production by 2050,’’ PLoS ONE, vol. 8, 1473 no. 6, 2013, Art. no. e66428. 1474

[4] G. N. Tiwari, Greenhouse Technology for Controlled Environment. 1475 Oxford, U.K.: Alpha Science Int.’l Ltd, 2003. 1476

[5] Historical Background of Greenhouses. Emerald Agr. Technol., 1477 Kolhapur, India. Accessed: May 1, 2021. [Online]. Available: 1478 http://www.emerald-agri.com 1479

[6] S. Vatari, A. Bakshi, and T. Thakur, ‘‘Green house by using IoT and 1480 cloud computing,’’ in Proc. IEEE Int.

Conf. Recent Trends Electron., Inf. 1481 Commun. Technol. (RTEICT), May 2016, pp. 246–250. 1482

[7] S. El-Gayar, A. Negm, and M. Abdrabbo, ‘‘Greenhouse operation and 1483 management in Egypt,’’ in Conventional Water Resources and Agricul- 1484 ture in Egypt. Cham, Switzerland: Springer, 2018, pp. 489–560. 1485

[8] I. L. López-Cruz, E. Fitz-Rodríguez, R. Salazar-Moreno, A. Rojano- 1486 Aguilar, and M. Kacira, ‘‘Development and analysis of dynamical math- 1487 ematical models of greenhouse climate: A review,’’ Eur. J. Hortic. Sci., 1488 vol. 83, pp. 269–280, Oct. 2018

[9] J. Xu, B. Gu, and G. Tian, ‘‘Review of agricultural IoT technology,’’ Artif. Intell. Agricult., vol. 6, pp. 10–22, Jan. 2022.

[10] (2022). World Economic Situation and Prospects As of Mid2022. Accessed: Jun. 14, 2023. [Online]. Available: www.un.org/ development/desa/dpad/publication/world- economic-situation-andprospects-as-of-mid-2022/

[11] (Sep. 2019). How to Feed the World in 2050 By Fao. Accessed: Jun. 14, 2023. [Online]. Available: https://www. fao.org/wsfs/forum2050/wsfs-forum/en/

[12] A. D. Tripathi, R. Mishra, K. K. Maurya, R. B. Singh, and D. W. Wilson, ‘‘Estimates for world population and global food availability for global health,’’ in The Role of Functional Food Security in Global Health. Amsterdam, The Netherlands: Elsevier, 2019, pp. 3–24.

[13] M. S. Farooq, S. Riaz, A. Abid, K. Abid, and M. A. Naeem, ‘‘A survey on the role of IoT in agriculture for the implementation of smart farming,’’ IEEE Access, vol. 7, pp. 156237–156271, 2019.

[14] M. Raj, S. Gupta, V. Chamola, A. Elhence, T. Garg, M. Atiquzzaman, and D. Niyato, ‘‘A survey on the role of Internet of Things for adopting and promoting agriculture 4.0,’’ J. Netw. Comput. Appl., vol. 187, Aug. 2021, Art. no. 103107.

[15] (Sep. 2019). Global Smart Agriculture Marke. Accessed: Jun, 14, 2023. [Online]. Available: https://www.zionmarketresearch.com/report/smartagriculture-market

[16] L. García, L. Parra, J. M. Jimenez, J. Lloret, and P. Lorenz, ‘‘IoT-based smart irrigation systems: An overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture,’’ Sensors, vol. 20, no. 4, p. 1042, Feb. 2020.

[17] W.-L. Chen, Y.-B. Lin, Y.-W. Lin, R. Chen, J.-K. Liao, F.-L. Ng, Y.-Y. Chan, Y.-C. Liu, C.-C. Wang, C.-H. Chiu, and T.-H. Yen, ‘‘AgriTalk: IoT for precision soil farming of turmeric cultivation,’’ IEEE Internet Things J., vol. 6, no. 3, pp. 5209–5223, Jun. 2019

. [18] J. Mendez and D. Dasig, ‘‘Frost prediction in highland crops management using IoT-enabled system and multiple regression,’’ in Internet of Things and Analytics for Agriculture, vol. 2. Cham, Switzerland: Springer, 2020, pp. 261–288.

[19] S. V. Gaikwad, A. D. Vibhute, K. V. Kale, and S. C. Mehrotra, ‘‘An innovative IoT based system for precision farming,’’ Comput. Electron. Agricult., vol. 187, Aug. 2021, Art. no. 106291

[20] R. Akhter and S. A. Sofi, ‘‘Precision agriculture using IoT data analytics 1524 and machine learning,’’ J. King Saud Univ.-Comput. Inf. Sci., vol. 34, 1525 no. 8, pp. 5602–5618, 2021. 1526

[21] K. Bakthavatchalam, B. Karthik, V. Thiruvengadam, S. Muthal, D. Jose, 1527 K. Kotecha, and V. Varadarajan, ‘‘IoT framework for measurement and 1528 precision agriculture: Predicting the crop using machine learning algo- 1529 rithms,’’ Technologies, vol. 10, no. 1, p. 13, 2022

[22] M. Pritts, L. McDermott, K. Demchak, E. Hanson, C. Weber, A. Both, and 1532 C. Heidenreich, ‘‘High tunnel raspberries and blackberries,’’ New York 1533 Fruit Quart., vol. 17, no. 3, pp. 13–16, 2009. 1534

[23] N. A. Farooqui, A. K. Mishra, and R. Mehra, ‘‘IOT based automated 1535 greenhouse using machine learning approach,’’ Int. J. Intell. Syst. Appl. 1536 Eng., vol. 10, no. 2, pp. 226–231, 2022. 1537

[24] (2022). Cochrane [Cochrane Collaboration]. Web. Accessed: 1538 Jul. 24, 2022. [Online]. Available: https://www.cochrane.org/ 1539

[25] Z. Q. R. Wang, Q. Chen, Y. Liu, and W. Qin, ‘‘IoT gateway: Bridging1540 wireless sensor networks into Internet of Things,’’ in Proc. IEEE/