

ENHANCING AGRICULTURAL PRODUCTIVITY THROUGH DATA SCIENCE AND MACHINE LEARNING

Anitha.P

Assistant Professor, Department of Computer Engineering
C Byregowda Institute of Technology, Kolar,
Karnataka, India.
anithacbit2@gmail.com

Dr.Ramesh Kumar.V

Professor and Head, Department of Computer Engineering
C Byregowda Institute of Technology, Kolar, Karnataka, India.
rameshkv.eee@gmail.com

Vaani Shree

Assistant Professor, Department of Computer Engineering
C Byregowda Institute of Technology, Kolar, Karnataka, India.
vanikasnoor.cbit@gmail.com

Abstract: The research utilizes methods such as data analysis, predictive modeling, and machine learning algorithms to examine agricultural data and generate actionable insights. This study focuses on optimizing agricultural practices through the application of data science and machine learning techniques. Agriculture plays a crucial role in ensuring global food security and economic stability, making its efficiency and sustainability a priority. Key objectives include enhancing crop productivity, reducing resource wastage, and improving overall farm management. The findings aim to support informed decision-making for farmers, policymakers, and stakeholders. Ultimately, this work contributes to the advancement of sustainable agriculture by promoting efficient resource utilization and increasing profitability

Keywords: Optimization, Data Analysis Agriculture Machine Learning, Data Science, Modeling, Crop Productivity.

I. INTRODUCTION

All This model is based on programming and involves computational methods. It follows several established principles that serve as tools for analyzing and simulating agricultural production strategies, evaluating the effects of different farming practices. As the population continues to grow, agriculture has become more sophisticated, and experienced farmers possess detailed knowledge about crop cultivation and the required growth intervals. Extensive variety of crops, including information such as soil humidity, pH levels, rainfall, temperature, and other relevant factors.

Farmers decide which crops to grow based on climate conditions, soil characteristics, and regional demand. For example, mild weather in spring makes it suitable for planting crops like tomatoes and cucumbers. During the hotter summer months, crops such as corn and melons are high temperatures. In the

cooler fall season, farmers grow and harvest crops like wheat and pumpkins, which adapt well to lower temperatures.

The most appropriate crop for given environmental conditions to achieve the highest possible yield and profitability. The structure of this work is organized as follows: Section 2 discusses existing agricultural practices based on previous research. Section 3 explains the data science process applied to agricultural yield. Section 4 focuses on data visualization techniques, followed by alternative farming methods in Section 5. In addition to seasonal planning, advanced farming methods rely on data and technology. Tools like sensors, weather automated systems and predictions, help farmers manage irrigation, control pests, and apply fertilizers effectively.

The smart practices improve productivity and crop quality while also reducing waste and protecting the environment. Modern agriculture has advanced significantly, adopting smarter and more efficient techniques. In today's world, data science is widely applied in many areas, including agriculture. It focuses on analyzing data to informed decision-making Important skills needed in this field include deep learning, data visualization, mathematics, and programming.

II. RELATED WORK

Agriculture is increasingly being transformed through digitization using artificial intelligence & the Internet of Things (IoT), moving from initial concepts to real-world applications. A control system built using a microcontroller regulates the flow of water and the application of fertilizers. Furthermore, a mobile application allows users to monitor and control the system remotely. Additional features like animal detection and crop protection are also integrated into the system.

Machine learning techniques, combined with specialized algorithms, enable accurate estimation of crop water requirements, ensuring precise and efficient irrigation. The application of Internet of Things (IoT) technology in automated farming systems significantly enhance the efficient use of water and fertilizers. These systems utilize sensors to continuously monitor key parameters such as humidity, temperature, and soil moisture levels.

The core technical aspects of AI and IoT, difficulties in adopting these technologies, have been explored in detail. By integrating digital technologies into farming practices, AI- and IoT-based solutions can be effectively implemented in real-world agricultural settings. These technologies enhanced efficiency, cost reduction, and improved crop productivity. However, more research and practical deployment are necessary to fully harness their capabilities in agriculture.

Recent developments in digital based solution for detecting and irrigating plants, agriculture have been reviewed, highlighting key applications in agricultural engineering. This review draws on research from major scientific databases such as PubMed, Web of Science, and Scopus, with a focus on studies published over the past decade. Cloud computing has gained significant attention in recent years. Overall, such systems play a vital role in advancing modern agriculture and contributing to the nation's economic growth. For data storage, analysis, and decision-making, cloud servers are widely used. Development tools such as embedded systems, MPLAB, and Proteus software support the implementation of these technologies.

The increasing global population and rising food demand have made agricultural automation a critical priority. Traditional farming methods are no longer sufficient to meet these needs and may contribute to soil degradation.

Advanced technologies such as IoT, wireless communication machine learning, convolutional neural networks (CNNs), artificial intelligence, and deep learning offer effective alternatives, as highlighted in [3].

Automation plays a key role in solving various agricultural issues such as leaf infections, crop diseases, storage problems, pesticide handling, weed control, and water management. Urgent attention is required to address challenges like the overuse of harmful chemicals, improper irrigation practices, environmental pollution, and their negative impacts on ecosystems. The adoption of automated systems helps in enhancing soil quality and improving crop productivity. The adoption of automated systems enhances soil quality and improves crop productivity.

This paper reviews the contributions researchers who have investigated different automation approaches in agriculture. Agriculture is being progressively reshaped by the adoption of digital technologies such as artificial intelligence and the Internet of Things (IoT), transitioning from theoretical concepts to practical implementations. The essential technical components of AI and IoT, along with the challenges associated with their adoption, have been carefully analyzed. Integrating these technologies into farming practices enables the effective deployment of AI- and IoT-based solutions in real-world agricultural settings.

The work presented in [3] proposes an IoT-Recent developments in digital based solution for detecting and irrigating plants including flowers and leaves, in botanical gardens. IoT technologies have shown significant success in streamlining and automating farming operations. Nevertheless, additional research and real-world implementation are needed to achieve more effective and reliable automation solutions in agriculture.

In paper [4], the authors discuss the process of implementing artificial intelligence agents across various sectors such as agriculture, finance, industry, and security. Machine learning techniques are used to analyze statistical data and previously collected information to carry out specific tasks. The growth of data science has been largely driven by these techniques. Modern applications are capable of executing complex functions, including analyzing historical data, collecting new data, facial recognition, and weather forecasting.

The study in [5] further explores the future scope and opportunities of automation and the Internet of Things (IoT) in agriculture. Moreover, smart agriculture utilizes sensors and advanced technologies to effectively manage soil moisture, temperature, and irrigation processes. The research in [5] investigates the future potential and opportunities of automation and the Internet of Things (IoT) in the agricultural sector.

Automation and robotics are playing an increasingly crucial role in agriculture, particularly in improving irrigation efficiency and reducing water wastage in areas such as horticulture, parks, gardens, and golf courses. Intelligent systems also enhance food safety by ensuring the correct selection and application of chemicals. Multi-tasking robots boost productivity by performing several operations quickly while maintaining consistent quality.

This review paper [6] presents development of a prototype smart farming system using technology, with the goal of providing cost-effective solutions for farmers. The system is integrated with sensors such as temperature, humidity, and water level sensors, and can be easily monitored and controlled. The research also focuses on developing and testing application software to ensure effective communication between system components. Technologies such as Node-RED, IBM Cloud, and IBM Watson are used to implement the system successfully. Overall, the paper offers valuable insights into IoT-based smart agriculture and its potential to improve modern farming practices.

III. DATA SCIENCE PROCESS IN AGRICULTURE YIELD

In this stage, the necessary data is identified and gathered from multiple sources to support model creation. These sources may include web server logs, social media data, census datasets, and real-time data streams obtained through APIs. Data plays a crucial role in developing a model for improving crop yield.

For this analysis, a dataset obtained from Kaggle is used, which contains 2200 rows and 8 columns. Each record in the dataset corresponds to a specific crop instance. The columns represent important soil and environmental attributes such as Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, rainfall, pH level, and the crop label.

IV. DATA VISUALIZATION

Histogram plots were used to visualize the distribution of crop characteristics.

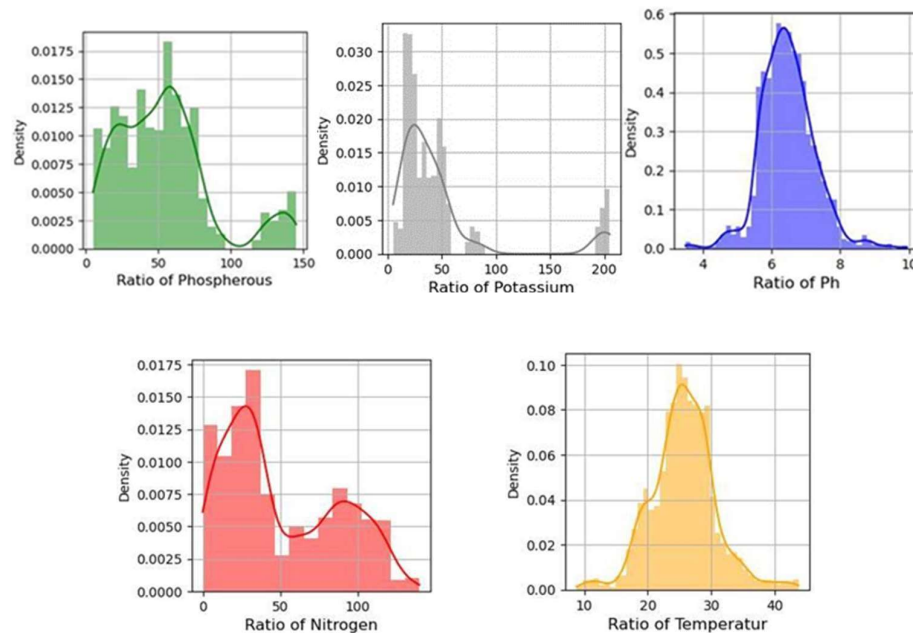


Fig. 1 Visualization of properties of crops

Patterns in Data

```
crops which requires very high ratio of Nitrogen  
content in soil: ['cotton']  
crops which requires very high ratio of Phosphorous  
content in soil: ['grapes' 'apple']  
crops which requires very high ratio of Potassium  
content in soil: ['grapes' 'apple']  
crops which requires very high Rainfall : ['rice']
```

Fig. 2 Crops that requires specific nutrient content for its growth in the soil.

Figure 1 presents a range of visualizations from the selected dataset, showcasing crop characteristics such as the levels of phosphorus, nitrogen, potassium, temperature, and pH.

The given analysis helps establish the relationship between soil conditions and crop requirements. This study is to develop a system that identifies the soil properties necessary for the optimal growth of specific crops, with a focus on key factors that affect crop productivity and health.

The research aims to develop an intelligent framework that processes collected data and applies algorithms to identify for specific crops, ultimately providing accurate and practical recommendations for farmers. As shown in Figure 2, key parameters such as nutrient composition, pH level, moisture retention, and soil structure.

The ultimate goal of this research endeavor is to equip the system with the ability to offer tailored and data-driven insights to individuals engaged in agriculture. By identifying the specific soil characteristics required for successful crop cultivation, the system aims to optimize agricultural practices, increase crop yields, and efficient agricultural management. Overall, the goal is to provide data-driven guidance that can improve agricultural efficiency, increase crop yield, and support sustainable farming practices.

The purpose of creating a sophisticated model that can assimilate and process a myriad of soil-related information, ultimately leading to accurate recommendations for farmers and cultivators. Furthermore the investigation strives to establish a robust framework where

the system can effectively interpret the gathered data and apply intelligent algorithms to determine the specific soil conditions most conducive to the healthy growth and development of the selected crop.

This study aims to develop a system capable of discerning the ideal crops for cultivation during distinct seasons. By doing so, the system's ability to recommend suitable crops aligns with the changing environmental conditions, potentially maximizing yield. This investigation seeks to optimize agricultural productivity by ensuring that the crops chosen for each season are well-suited to prevailing weather patterns, soil characteristics, and other contextual factors, ultimately contributing to more effective and sustainable cultivation practices. Investigation done in this aspect is shown in figure 3.

Summer Crops
['pigeonpeas' 'mothbeans' 'blackgram' 'mango' 'grapes' 'orange' 'papaya']

Winter Crops
['maize' 'pigeonpeas' 'lentil' 'pomegranate' 'grapes' 'orange']

Rainy Crops
['rice' 'papaya' 'coconut']

Fig 3. Crops that grow in specific season

I. Alternate Farming Using Clusters

Alternate farming through clustering involves cultivating different but related crops in a rotational manner within the same area. This approach promotes soil fertility, lowers the risk of pests and diseases, and can lead to higher crop productivity. It also assists in identifying compatible crops that can be grown together to improve overall efficiency.

In summary, clustering-based alternate farming offers a sustainable solution for agriculture by enhancing crop planning, optimizing resource use, and reducing environmental impact [7]. Alternate farming through clustering involves cultivating different but related crops in a rotational manner within the same area.

A) Elbow method in cluster analysis

The elbow method is a commonly used technique in data science and machine learning to determine the optimal number of clusters in a dataset during cluster analysis. This point, known as the “elbow,” indicates the most appropriate number of clusters.

Selecting this point helps achieve a balance between model accuracy and complexity, avoiding both overfitting and underfitting. The “elbow” represents the stage where the reduction in variance begins to level off, forming a bend in the curve. This point is considered ideal as it provides a good trade-off between fitting the data effectively and maintaining simplicity. Figure 4 illustrates the relationship between the number of clusters and crops.

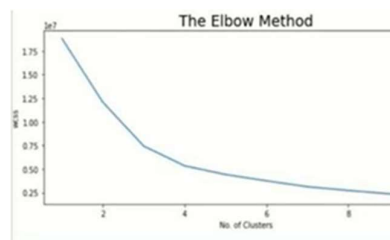


Fig 4. Cluster identification using elbow method.

B) K-Means Algorithm

Clustering can be carried out using several approaches, such as graph-based, grid-based, model-based, hierarchical, partitioning, and density-based methods. These six categories broadly define the different types of clustering

techniques. Among them, K-means is one of the most commonly used algorithms and falls under the partitioning method.

In the K-means algorithm, the user specifies the number of clusters (K), and the dataset containing N data points is divided into K groups. The goal is to ensure that data points within the same cluster are highly similar (high intra-cluster similarity), while those in different clusters are as dissimilar as possible (low inter-cluster similarity).

In this study, the dataset consisting of N data points is partitioned into K clusters using a clustering technique that follows these steps:

- 1.) K data points are randomly selected from the dataset (DS) to act as initial cluster centers (CC).
- 2.) Each data point is assigned to the cluster with the nearest center based on similarity measures.
- 3.) The cluster centers are updated by computing the mean of all data points within each cluster, and these mean values become the new centers.
- 4.) Steps 2 and 3 are repeated iteratively until the cluster assignments stabilize and no further changes occur.

After Applying KMeans For Clustering Analysis

Crops in First Cluster : ['m

Crops in Second Cluster : ['maize' 'aize' 'chickpea' 'kidneybeans' 'pigeonpeas' 'mothbeans' 'mungbean'

'blackgram' 'lentil' 'pomegranate' 'mango' 'orange' 'papaya' 'coconut']

- 'banana' 'watermelon' 'muskmelon'

'papaya' 'cotton' 'coffee']

Crops in Third Cluster : ['grapes' 'apple']

Figure 5 shows the grouping of similar crops within each cluster, and Figure 6 illustrates the K-means clustering procedure.

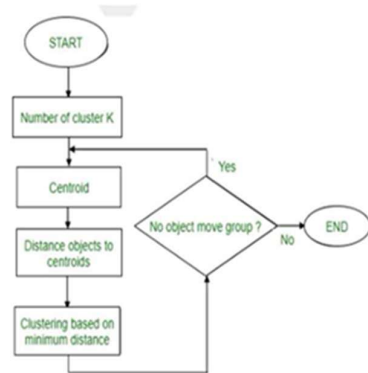


Fig 6. Flowchart representing the Methodology in K- Means

VI. Predictive Model

This section outlines the predictive models examined in this study. Regression is a statistical technique used to analyze relationships between variables and make predictions. Logistic regression is one such method,

commonly used in predictive modeling due to its simplicity and interpretability. It estimates the probability of a specific outcome based on a set of input features.

In logistic regression, the dependent variable is binary, meaning it has only two possible outcomes (such as 0 or 1). The objective is to model the probability that the outcome equals 1 given the predictor variables. These predictors may be continuous or categorical and help determine the likelihood of the event.

Once trained, the logistic regression model can be used to make predictions on new data. It produces probability scores, which can be converted into binary outcomes (such as yes or no) using a predefined threshold be used to make predictions on new data.

VII. CONCLUSION

As the global population continues to grow rapidly, it is important to implement effective and forward-looking measures. Agricultural optimization serves as a key approach to determine the most suitable crops based on soil and environmental conditions. This study considers essential factors that influence plant growth, including rainfall, temperature, soil nitrogen, potassium, phosphorus, humidity, and pH level, as these play a vital role in crop development.

In conclusion, this method can support farmers in evaluating soil and weather conditions to choose the most appropriate crops. By optimizing agricultural practices, it enables sustainable farming, enhances productivity, and ensures efficient use of agricultural land.

REFERENCES

- [1]. S.A.Sivakumar,G.Mohanapriya, A.Rashini, R.Vignesh," Agriculture Automation using Internet of Things", International Journal of Advance Engineering and Research Development, Vol.5, No.2, Feb 2018
- [2]. No.2, Feb 2018
- [3]. A. Subeesh, C.R. Mehta," Automation and digitization of agriculture using artificial intelligence and internet of things", Artificial Intelligence in Agriculture, Vol.5, 2021.
- [4]. Kirtan Jha, Aalap Doshi, Poojan Patel, Manan Shah," A comprehensive review on automation in agriculture using artificial intelligence", Artificial Intelligence in Agriculture, Vol.2, 2019
- [5]. G. Ramkumar, M. Alluwaici, T. Vijetha, G. H, A. Venkatraman and R. M, "A Reliable System Performance Through NeuToMa with Spiking Neural Network Transformation in Network on Chip (NoC) Communication," 2026 International Conference on Communication, Computing and Emerging Technologies (IC3ET), Vasai, India, 2026, pp. 215-219, doi: 10.1109/IC3ET64989.2026.11467379.
- [6]. V. R, P. R. Kumar, D. Das, G. H, W. Hanbin and E. A, "Machine Learning-Driven Predictive Maintenance for Wind Turbines," 2026 International Conference on Electric Power and Renewable Energy (EPREC), Durg, India, 2026, pp. 1-6, doi: 10.1109/EPREC66546.2026.11412064.
- [7]. A. Tiwari, J. D. Kumar, W. Syed and G. H, "Efficient FPGA-Based CNN Accelerator for Real-Time Traffic Sign Classification on PYNQ-Z2," 2026 IEEE International Conference on AI Engineering and Innovations (AIEI), NIT Jamshedpur, India, 2026, pp. 1-4, doi: 10.1109/AIEI69164.2026.11496958.

- [8]. G. Ramkumar, M. Alluwaici, T. Vijetha, G. H, A. Venkatraman and R. M, "A Reliable System Performance Through NeuToMa with Spiking Neural Network Transformation in Network on Chip (NoC) Communication," *2026 International Conference on Communication, Computing and Emerging Technologies (IC3ET)*, Vasai, India, 2026, pp. 215-219, doi: 10.1109/IC3ET64989.2026.11467379.
- [9]. Girish H., Srinivas S., Kavya M., ...Kabbur S. "Swift Autonomous Fire Suppression Robot", 16th International Conference on Advances in Computing Control and Telecommunication Technologies Act 2025
- [10]. Girish H., Soujanya S., "A Comprehensive Cadence-based Design Flow for RISC-V UART SOC IN SCL 180NM", 16th International Conference on Advances in Computing Control and Telecommunication Technologies Act 2025
- [11]. Girish H., Rakesh S., Prasanna Raj C.P., Singh A.S., Yathisha L, "Deep Learning for Diabetic Retinopathy: Automated Detection and Classification from Retinal Images", 16th International Conference on Advances in Computing Control and Telecommunication Technologies Act 2025
- [12]. Girish H., Shyam Sundar V., Priyanka C., Renuka Hebballi R.S, "RISC V Steel Processor Implementation for Communication between Two Chips using AXI Lite", 16th International Conference on Advances in Computing Control and Telecommunication Technologies Act 2025
- [13]. M. H N, G. H and N. K N, "Simulation and Verification of SPI Master and Slave Communication," *2025 6th International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, 2025, pp. 1531-1534, doi: 10.1109/ICOSEC67334.2025.11459684. <https://ieeexplore.ieee.org/document/11459684>
- [14]. S. S. V, P. C and G. H, "Design and Analysis of a D-Band LC Bandpass Filter and Low-Noise Amplifier for 6G Front-End Applications," *2026 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India, 2026, pp. 1-6, doi: 10.1109/IATMSI68868.2026.11466038. <https://ieeexplore.ieee.org/document/11466038>
- [15]. "A Comprehensive Cadence-based Design Flow for RISC-V UART SOC IN SCL 180NM" , Girish H, Soujanya, Grenze International Journal of Engineering and Technology, 2025
- [16]. "Swift Autonomous Fire Suppression Robot" Girish H et. Al., Grenze International Journal of Engineering and Technology, 2025
- [17]. "Deep Learning for Diabetic Retinopathy: Automated Detection and Classification from Retinal Images" , Girish H, Cyril Prasanna Raj P , Grenze International Journal of Engineering and Technology, 2025
- [18]. "RISC V Steel Processor Implementation for Communication between Two Chips using AXI Lite" Shyam Sundar V, Dr Girish H, Priyanka C, Renuka S Hebballi Grenze International Journal of Engineering and Technology, 2025
- [19]. Sujatha.S, Mahendra BM, & Dr. Girish H. (2024). Smart Irrigation System for Optimal Water Conservation. *Revista Electronica De Veterinaria*, 25(1), 2613 - 2619. <https://doi.org/10.69980/redvet.v25i1.1311>
- [20]. Mamatha K R, Thejaswini S, Rashmi N, & Dr. Girish H. (2024). Laser Fencing Surveillance System With Email Alert. *Educational Administration: Theory and Practice*, 30(2), 1539–1544. <https://doi.org/10.53555/kuey.v30i2.7919>
- [21]. Thejaswini S, Mamatha K R, Rashmi N, & Dr. Girish H. (2024). Performance Evaluation of AI Models in Early Heart Disease

- Diagnosis. Educational Administration: Theory and Practice, 30(2), 1531–1538. <https://doi.org/10.53555/kuey.v30i2.7918>
- [22]. Sushma K Sattigeri, Seema Srinivas, Shridhar Kabbur, Girish H, “Revolutionizing Urban Waste Management: An AI-Powered Autonomous Solution for Smart Cities”, *Tuijin Jishu/Journal of Propulsion Technology* ISSN: 1001-4055, Vol. 45 No. 02 (2024)
- [23]. Girish H, Vajjala Sai Sree Hari, Ponde Lakshmi Narasimha, P N Shreya, Dadigala Bhargavi, "THE DESIGN AND VERIFICATION OF A 32-BIT RISC V PROCESSOR USING VEDIC MATHEMATICS", *International Journal of Creative Research Thoughts (IJCRT)*, ISSN:2320-2882, Volume.12, Issue 5, pp.1130-1135, May 2024, Available at :<http://www.ijcrt.org/papers/IJCRTAB02172.pdf>
- [24]. IoT based home automation using Blynk cloud, J. Chetan Naik, B. S. Deeksha, H. Girish, Gurusharan Kashetty, T. G. Shivapanchakshari, Manohar Sai Polakam, *Journal: AIP Conference Proceedings*, AIP Conf. Proc. 2965, 020006 (2024), <https://doi.org/10.1063/5.0215427>, Published: July 2024
- [25]. IoT based fire alarm system using Blynk app, T. Bhargava, H. Kishor Kumar, Pramod Mahalingappa Godi, H. Girish, J. Chetan Naik, *Journal: AIP Conference Proceedings*, AIP Conf. Proc. 2965, 020007 (2024), <https://doi.org/10.1063/5.0212325>, Published: July 2024