Grenze International Journal of Engineering and Technology, Jan Issue

Regression-based Smart Agriculture Prediction

Dr. Sachin Kumar[1], Mr.Hirdesh Sharma[2], Dr. Sheetal Rajput[3] ,Dr. Arun Kumar[4] and Mr.Vijay Kumar Tiwari[5]
[1]Associate Professor, Dept. of Information Technology, Management Education and Research Institute (MERI), Affiliated to GGSIP University, New Delhi, India,
ORCID ID: 0000-0002-1136-8009, Email: sachinks.78@gmail.com
[2]Assistant Professor, Dept. of CSE, JIMS Engineering Management Technical Campus, (JEMTEC), Greater Noida, U.P., India, ORCID ID: 0000-0002-1278-4135
Email: hirdesharma@gmail.com
[3]Associate Professor in Dept. of Computer Science, Bhagwant Institute of Technology,Muzaffarnagar, U.P, India, ORCID ID: 0009-0001-4147-788X
Email: sheetal1.rajput@gmail.com
[4]Assistant Professor, Dept. of CSE, İTS Engineering College, Greater Noida, U.P., India, ORCID ID: 0000-0002-5493-8059
Email: scorearun84@gmail.com
[5]Assistant Professor, Dept. of CSE, İTS Engineering College, Greater Noida, U.P., India, ORCID ID: 0000-0002-5771-8975
Email: vijayvijay456@gmail.com

***Abstract*— Smart agriculture has come to be seen as a potential method for boosting agricultural output and sustainability. The study primarily focuses on the application of regression algorithms for predictive analysis in smart agriculture with reference to crop yield prediction. This study's main objective is to develop a reliable regression model that estimates crop yields based on a number of agricultural and environmental factors. This is done by accumulating previous agricultural information, such as weather patterns, soil properties, irrigation methods, and crop-specific traits, via smart sensors positioned throughout agricultural areas. In the study, a variety of regression approaches are investigated, including support vector regression (SVR), random forest regression, linear regression, polynomial regression, and more. The dataset is analyzed before using these methods to build prediction models. To validate the model's accuracy and generalizability, the dataset is split into training and testing sets, and cross-validation techniques are employed. Regression algorithms are highly reliable at predicting agricultural yields, according to the results. Random forest regression outperforms all other methods due to its capability to handle non-linearity and capture nuanced relationships between input characteristics and crop yields. However, the findings from other regression algorithms are equally promising, suggesting that they may be useful in specific situations. Additionally, the study evaluates how different input variables impact the model's efficacy. It draws attention to crucial elements including temperature, precipitation, soil moisture, and nutrient levels that have a significant impact on crop yields. Farmers may select the ideal crops, irrigation schedules, and fertilizer applications with the aid of an understanding of these connections.**

***Index Terms*— Smart agriculture, Machine Learning, Regression Algorithm.**

I. INTRODUCTION

Concern over the need to use resources responsibly has grown recently, especially in the context of agriculture, which is crucial for feeding the world's expanding population. Modern technology is being used to expedite

Grenze ID: 01.GIJET.10.1.522_2

farming operations, boost output, and lessen the impact on the environment, replacing the outdated agricultural practices with "Smart Agriculture," a creative approach. The quickening pace of technological advancement is driving this change. Predictive analytics, one of the core components of Smart Agriculture, relies heavily on regression algorithms to generate accurate forecasts and sensible recommendations. Smart agriculture seeks to enhance a variety of farming practices, including irrigation control, crop yield forecasts, disease detection, and resource allocation, by utilizing data-driven insights and prediction models. Regression algorithms, a subset of machine learning techniques, are highly helpful in this industry because they can examine historical data, identify trends, and create connections between variables to predict future outcomes. By employing regression models, farmers can gain valuable insights about crucial agricultural characteristics, allowing them to take preemptive action and change their approach as needed [1-2].

The key objectives of using regression algorithms in smart agriculture are as follows:

(1) *Yield Prediction*: By reviewing historical data on weather patterns, soil properties, irrigation schedules, and crop kinds, regression algorithms can generate accurate estimates of agricultural yields for various seasons and locations.

(2) *Resource Optimization*: Predictive models can also aid with timing and dosage of fertilizer application, reducing waste and unfavourable environmental effects.

(3) Pest and Disease Detection: Regression algorithms can be trained to forecast potential outbreaks and suggest acceptable preventive measures or treatment techniques using data on pest occurrences, weather trends, and crop health indicators.

(4) *Climate Resilience*: Due to climate change, agriculture will confront numerous challenges. Farmers can use regression models to look at past weather data and identify climatic trends and patterns. Planning for crops that are climate resilient, altering planting dates, and utilizing adaptive agriculture practices are all effective ways to mitigate the consequences of climate change.

(5) *Market Trends and Prices*: The analysis of historical market data, such as changes in agricultural demand and prices, can also be done using regression methods. The capacity to foresee market trends, choose crops wisely, and increase revenue is provided to farmers as a result.

Regression algorithms could therefore fundamentally change how agriculture is now practiced. As technology advances, the integration of regression algorithms into agriculture will be essential for ensuring the economic success, environmental sustainability, and food access of farming communities around the world [3-4].

II. Application Of Machine Learning In Agriculture

In order to help farmers, make better decisions and boost productivity, machine learning algorithms can reliably predict crop yields, weather patterns, disease outbreaks, irrigation plans, and more by analysing historical and current data [5].

For agricultural prediction, the following crucial machine learning components are included:

Crop Yield Prediction: This information will enable farmers to make better resource allocation choices, planting strategy plans, and decision-making that will boost crop output [6].

- *Disease Detection and Prevention*: The relationship between environmental factors and crop disease outbreaks can be detected by machine learning models as trends and links.
- *Irrigation management*: By studying weather patterns, crop water requirements, and soil moisture data, machine learning algorithms can optimize irrigation schedules and water use. Water resources are preserved and over- or under-irrigation is prevented by doing this.
- *Control of Pests and Weeds*: By analysing sensor data or drone photos, machine learning can support the monitoring of weed and pest populations. Machine learning algorithms can identify patterns and links between changes in the environment and agricultural disease outbreaks. Farmers can quickly and effectively treat diseases when they are identified early, reducing crop losses and the need for pesticides [8-9].
- *Irrigation management*: By studying information on soil moisture, weather patterns, and crop water needs, machine learning algorithms can improve irrigation schedules and water usage.
- *Control of Pest and Weed Populations*: Machine learning can help in monitoring pest and weed populations by looking at sensor data or drone image data.

III Why Smart Agriculture Prediction

Numerous convincing arguments have been made that smart agricultural prediction is crucial to modern farming techniques [11-12]:

Enhanced Efficiency: By accurately anticipating factors like crop yields, weather patterns, and pest outbreaks, farmers may better plan their activities, allocate resources, and minimize waste.
Enhanced Productivity: With accurate projections, farmers may make informed decisions about crop selection, planting dates, and other key issues.
Risk reduction: Farmers can proactively manage risks like disease outbreaks and extreme weather thanks to the use of predictive analytics in agriculture.
Technology adoption: Innovative methods like AI, machine learning, and IoT are used in smart agriculture prediction.
Global Food need: Smart agriculture prediction can help to fill this expanding requirement by maximizing agricultural production and guaranteeing that there is enough food for everyone on the earth. By categorizing the parameters according to the features, farmers are better able to predict which crop can be cultivated in a certain parameter [16].

IV. Parameters That Will Be Used In Smart Agriculture

Smart agriculture makes use of various metrics and data sources to enhance farming practices and productivity. Smart agriculture uses a variety of essential factors, including [17-18]:

- *Weather Information*: Understanding the climatic conditions that affect crop development and irrigation needs is crucial, thus it's vital to take into account meteorological factors like temperature, humidity, rainfall, wind speed, and sun radiation.
- *Crop Growth and Health*: By keeping an eye on the growth phases and general health of their crops using visual evaluation or remote sensing technology, farmers can see early signs of stress, illness, or pests and take appropriate action [19].
- *Irrigation data:* When irrigation practices are optimized, water is saved and crops develop to their full potential. This is made possible by knowledge of irrigation schedules, water usage, and soil moisture levels.
- *Pest and Disease Data*: Farmers are assisted in making crop decisions and creating marketing strategies by data on agricultural prices, market demand, and trends.
- Monitoring environmental factors including air quality, pollution levels, and carbon dioxide concentration can help shed light on how agricultural practices affect the environment as a whole. The management of the welfare and productivity of animals in animal agriculture depends heavily on the observation of the animals.
- *Data from equipment and machinery*: Monitoring data from farm machinery and equipment is another component of smart agriculture that helps to ensure proper maintenance and operation. By combining and assessing these many aspects, intelligent agriculture systems may provide farmers with insightful data.
- *Crop Type*: Crop type is a crucial characteristic used in smart agriculture to categories and separate various crops grown on a farm or in a specific area. Understanding the type of crop being grown is essential for managing a farm and making decisions because it has a direct impact on many agricultural practices and choices.
- *Crop Production*: Crop production involves planting, cultivating, and harvesting crops for use as food, animal feed, or industrial materials. It is a crucial component of agriculture and is critical for providing food, raw materials, and fiber to support multiple industries as well as human existence.
- *Geographical Area*: Geographical regions can range in size and extent from a small piece of land to a huge country or continent. A major determinant of many agricultural practices and decision-making processes in the context of smart agriculture is geographic area.
- *Rain Fall Amount*: Rainfall, often known as precipitation, is a critical environmental element in agriculture that has a significant impact on how well crops grow, how much water is available, and how efficiently farms are run as a whole.
- *Organic Fertilizer and Organic Compounds*: The basis of organic farming is organic fertilizer and organic substances, a method of agricultural production that shuns the use of synthetic chemicals and promotes ecological sustainability and balance.
- *Soil Quality*: Soil quality, or the general health and fertility of the soil, is a significant element in agriculture.

V. How We Predict

Utilizing the facts and relevant information that are currently available is necessary in order to make accurate predictions or guesses regarding prospective occurrences or outcomes. The following stages are frequently included in prediction [24-25]:

- *Data Collection*: A prediction may only be made after gathering relevant data from several sources. The information used only needs to be pertinent to the prediction task; examples include historical data, observations, sensor readings, survey findings, etc.
- *Data Preprocessing*: After the data has been collected, it is essential to preprocess and clean the data to remove noise, outliers, and missing values. Another component of data preparation is converting the data into an analytically usable format.
- *Feature Selection and Engineering*: Predictions will be made using the attributes (features) from the data that are most relevant and instructive. Feature engineering, which entails creating fresh features or representations, is another method for raising the predictive capability of the model.
- *Model Selection*: It's critical to choose the best prediction model. The type of prediction job being used (such as regression, classification, or time-series forecasting) as well as the characteristics of the data affect the model selection.
- *Model Evaluation*: To assess accuracy, precision, recall, or other relevant metrics, several metrics are used depending on the details of the prediction task.
- *Prediction and Forecasting*: After being trained and validated, the model can be utilized to make predictions on recent, undiscovered data. As a time-series forecasting tool, the model may predict future values by using historical data.
- *Continuous Improvement*: The process of making predictions is iterative and ongoing.

The use of prediction techniques is widespread in many different fields, including finance, medicine, weather forecasting, agriculture, and many more.

VI. K- Nearest Neighbors Regression Analysis

It is possible to get the average of the numerical target of the K nearest neighbours using a simple KNN regression implementation. One such technique uses the inverse distance-weighted average of the K nearest neighbours. KNN regression uses identical distance function as KNN classification.

Euclidean $$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

Manhattan $$\sum_{i=1}^{k}|x_i - y_i|$$

Minkowski $$\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$$

Fig 1 Distance Functions

$$D_H = \sum_{i=1}^{k}|x_i - y_i|$$

$$x = y \Rightarrow D = 0$$

$$x \neq y \Rightarrow D = 1$$

Fig 2 Hamming Distance

Only continuous variables are acceptable for the three distance metrics mentioned above. The Hamming distance, a measurement of the number of times equivalent symbols in two strings of equal length are different, must be used when dealing with categorical variables. It is best to examine the data first before selecting the best value for K. Using a different data set to validate your K value is a technique known as cross-validation. It is another way to determine an appropriate K value in the past. For most datasets, a value of K greater than 10 is ideal.

VII. Results

This dataset is used to suggest a crop for a certain soil type. This will be extremely helpful in crop production in (Smart Agriculture) without losses dependent on soil ph., rainfall, humidity, and other chemical elements present in the soil. Df is equal to pd. read_csv ("../input/smart-agricultural-production-optimizing engine/Crop_recommendation.csv").

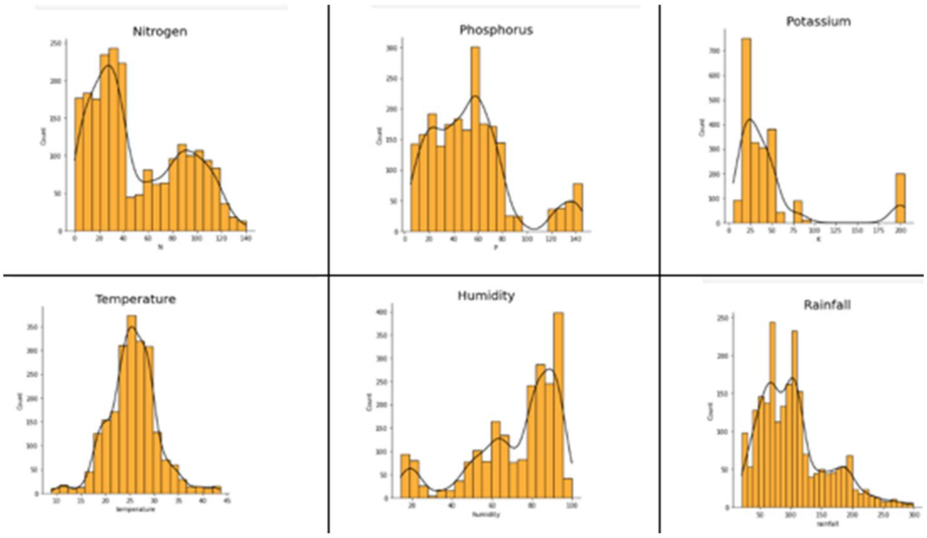

Fig 3: Prediction of smart agriculture on different parameters

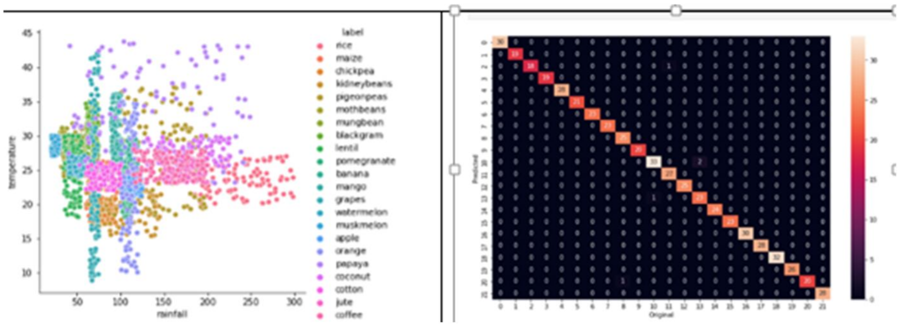

Fig 4. Prediction of various fruits and vegetables in smart agriculture and its heatmap

VIII. Challenges Of Smart Agriculture Prediction

The implementation and widespread use of smart agriculture prediction must overcome a number of challenges, despite the fact that it is advantageous and promising. Particularly in isolated or underdeveloped areas, it might be challenging to ensure the high degree of dependability, accuracy, and accessibility of this data. It might be difficult to integrate and use data that comes from numerous sources and is presented in a variety of formats.

Connectivity and Infrastructure: Due to a lack of reliable internet connectivity and infrastructure, real-time data transfer for analysis and forecasting may be challenging in many agricultural settings.

Limited Adoption and Awareness: The adoption of predictive models and smart agriculture technology may be slowed down by farmers who are unaware of them or who are unwilling to use them.

Cost and affordability: Having limited access to these sophisticated instruments and complex environmental factors, small-scale farmers may not be able to afford the initial investment required for smart agricultural equipment and predictive models. When these factors are taken into consideration, developing predictive models may be challenging.

Data security and privacy: Data about agriculture acquired from a variety of sources may contain sensitive information about farmers' practices and properties. Ensuring data security and privacy is necessary to win the trust of farmers and stakeholders alike.

Lack of Technical Knowledge: It's likely that agricultural stakeholders and farmers lack the technical knowledge necessary to understand, assess, and use the findings of predictive models for smart agriculture.

Regulation and policy issues In order for smart agricultural forecast to function, regulations or laws on data sharing, environmental impact, and privacy may need to be adhered to internet capacity and connectivity issues: There are connectivity and internet bandwidth requirements for the sensor networks used in smart agriculture.

Managing Data Volume: High-speed internet access is necessary for smart agriculture, but occasionally it may not be available if it is set up on a vast volume of data.

The Steep Learning Curve (Lack of Technical skills): The farmer operating the sophisticated agricultural software lacked technical expertise and was not very qualified.

Farmers must be trustworthy and dependable if they are to use predictive models for decision-making. In order for smart agriculture prediction to be successful and realize its full potential, it is crucial to make investments in agricultural infrastructure, promote data standardization, and provide farmers with training and support.

IX. CONCLUSION

According to this study, regression algorithms can be used to accurately predict crop yields in smart agriculture. The recommended models can aid farmers in more efficient resource management, improved crop output, and ultimately the promotion of sustainable agricultural practices. The combination of cutting-edge data analytics and machine learning techniques into smart agriculture has the potential to totally alter the agricultural sector and provide food security for a growing global population as it advances. The suggested method makes use of various variables to determine the potential area's most productive crops, assisting the farmer in making the most profitable choice. This method can be developed further in the future using the IOT to get real-time values for the soil and other properties. The systems can increase the results' accuracy and precision while offering farmers a user-friendly online application or mobile application by putting sensors on a farm to collect data about a parameter. As a result, farming can be carried out shrewdly.

REFERENCES

[1] Shah, A., Dubey, A., Hemnani, V., Gala, D., & Kalbande, D. R. (2018). Smart farming system: Crop yield prediction using regression techniques. In Proceedings of International Conference on Wireless Communication: ICWiCom 2017 (pp. 49-56). Springer Singapore.

[2] Ruß, G., & Brenning, A. (2010). Data mining in precision agriculture: management of spatial information. In Computational Intelligence for Knowledge-Based Systems Design: 13th International Conference on Information Processing and Management of Uncertainty, IPMU 2010, Dortmund, Germany, June 28-July 2, 2010. Proceedings 13 (pp. 350-359). Springer Berlin Heidelberg.

[3] Bendre, M. R., Thool, R. C., & Thool, V. R. (2015, September). Big data in precision agriculture: Weather forecasting for future farming. In 2015 1st international conference on next generation computing technologies (NGCT) (pp. 744-750). IEEE.

[4] Gamage, R., Rajapaksa, H., Sangeeth, A., Hemachandra, G., Wijekoon, J., & Nawinna, D. (2021, October). Smart agriculture prediction system for vegetables grown in Sri Lanka. In 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 0246-0251). IEEE.

[5] Ruß, G., & Kruse, R. (2010). Regression models for spatial data: An example from precision agriculture. In Advances in Data Mining. Applications and Theoretical Aspects: 10th Industrial Conference, ICDM 2010, Berlin, Germany, July 12-14, 2010. Proceedings 10 (pp. 450-463). Springer Berlin Heidelberg.

[6] Arumugam, K., Swathi, Y., Sanchez, D. T., Mustafa, M., Phoemchalard, C., Phasinam, K., & Okoronkwo, E. (2022). Towards applicability of machine learning techniques in agriculture and energy sector. Materials Today: Proceedings, 51, 2260-2263.

[7] Mohapatra, A. G., & Lenka, S. K. (2016). Neuro-fuzzy-based smart DSS for crop specific irrigation control and SMS notification generation for precision agriculture. International Journal of Convergence Computing, 2(1), 3-22.

[8] Apat, S. K., Mishra, J., Raju, K. S., & Padhy, N. (2022). The robust and efficient Machine learning model for smart farming decisions and allied intelligent agriculture decisions. Journal of Integrated Science and Technology, 10(2), 139-155.

[9] Wei, M. C. F., Maldaner, L. F., Ottoni, P. M. N., & Molin, J. P. (2020). Carrot yield mapping: A precision agriculture approach based on machine learning. Ai, 1(2), 229-241.

[10] Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and electronics in agriculture, 151, 61-69.
[11] Abdallah, E. B., Grati, R., & Boukadi, K. (2022, June). A machine learning-based approach for smart agriculture via stacking-based ensemble learning and feature selection methods. In 2022 18th International Conference on Intelligent Environments (IE) (pp. 1-8). IEEE.
[12] Rokade, A., Singh, M., Malik, P. K., Singh, R., & Alsuwian, T. (2022). Intelligent data analytics framework for precision farming using iot and regressor machine learning algorithms. Applied Sciences, 12(19), 9992.
[13] Kwaghtyo, D. K., & Eke, C. I. (2023). Smart farming prediction models for precision agriculture: a comprehensive survey. Artificial Intelligence Review, 56(6), 5729-5772.
[14] Khan, A. A., Shaikh, Z. A., Belinskaja, L., Baitenova, L., Vlasova, Y., Gerzelieva, Z., ... & Barykin, S. (2022). A blockchain and metaheuristic-enabled distributed architecture for smart agricultural analysis and ledger preservation solution: A collaborative approach. Applied Sciences, 12(3), 1487.
[15] Chakrabarty, A., Mansoor, N., Uddin, M. I., Al-Adaileh, M. H., Alsharif, N., & Alsaade, F. W. (2021). Prediction approaches for smart cultivation: a comparative study. Complexity, 2021, 1-16.
[16] Singh, A., Janu, N., Trivedi, S., & Jain, M. (2022, June). Precision agriculture and machine learning. In 2022 IEEE World Conference on Applied Intelligence and Computing (AIC) (pp. 659-664). IEEE.
[17] Varghese, R., & Sharma, S. (2018, June). Affordable smart farming using IoT and machine learning. In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 645-650). IEEE.
[18] Kunapuli, S. S., Rueda-Ayala, V., Benavidez-Gutierrez, G., Córdova-Cruzatty, A., Cabrera, A., Fernandez, C., & Maiguashca, J. (2015). Yield prediction for precision territorial management in maize using spectral data. In Precision agriculture'15 (pp. 344-358). Wageningen Academic Publishers.
[19] Ahmed, M. S., Tazwar, M. T., Khan, H., Roy, S., Iqbal, J., Rabiul Alam, M. G., ... & Hassan, M. M. (2022). Yield Response of Different Rice Ecotypes to Meteorological, Agro-Chemical, and Soil Physiographic Factors for Interpretable Precision Agriculture Using Extreme Gradient Boosting and Support Vector Regression. Complexity, 2022, 1-20.
[20] Dimov, D., Uhl, J. H., Löw, F., & Seboka, G. N. (2022). Sugarcane yield estimation through remote sensing time series and phenology metrics. Smart Agricultural Technology, 2, 100046.
[21] Kumar, C. K., & Venkatesh, V. (2019). Design and development of IOT based intelligent agriculture management system in greenhouse environment. International Journal of Engineering and Advanced Technology, 8, 47-52.
[22] Elashmawy, R., & Uysal, I. (2023). Precision Agriculture Using Soil Sensor Driven Machine Learning for Smart Strawberry Production. Sensors, 23(4), 2247.
[23] Upadhyay, D., Dubey, A. K., & Thilagam, P. S. (2018). Application of non-linear gaussian regression-based adaptive clock synchronization technique for wireless sensor network in agriculture. IEEE Sensors Journal, 18(10), 4328-4335.
[24] Lesch, S. M., Wallender, W., & Tanji, K. (2012). Statistical models for the prediction of field-scale and spatial salinity patterns from soil conductivity survey data. Agricultural salinity assessment and management, 461-482.