

Agrarian Synthesis and Precision Cultivation Optimization System

Tharaniya S¹, Vignesh J², Nandhitha Karthikeyini M³,

Nijandhan K⁴

¹UG Scholar, Department of Electronics and Instrumentation Engineering, Bannari Amman Institute of Technology, Erode, India.

²UG Scholar, Department of Mechatronics Engineering, Bannari Amman Institute of Technology, Erode, India.

³UG Scholar, Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Erode, India.

⁴Associate Professor, Department of Mechatronics Engineering, Bannari Amman Institute of Technology, Erode, India.

E-mail: ¹tharaniya.ei20@bitsathy.ac.in, ²vigneshj.mc20@bitsathy.ac.in, ³nandhithakarthikeyini.cs20@bitsathy.ac.in, ⁴nijandhan@bitsathy.ac.in

Abstract

The ever-growing demand for food production calls for innovative solutions in agriculture. This research introduces a machine learning-based approach, specifically utilizing logistic regression, to predict optimal crops based on soil and weather conditions. The dataset encompasses crucial attributes including Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH, rainfall, with corresponding crop labels. The proposed methodology employs logistic regression, a powerful classification algorithm, to model the relationships between input features and crop types. Through careful feature engineering, the model is fine-tuned to enhance its predictive accuracy. Rigorous evaluation metrics validate the model's performance, ensuring its reliability in real-world applications. Results showcase the logistic regression model's efficacy in accurately predicting suitable crops for given soil and weather parameters. This predictive tool serves as a practical decision support system for farmers, aiding in crop selection and resource allocation. This research contributes to the synergy of machine learning and agriculture, showcasing logistic regression as a valuable tool

for crop prediction and resource optimization. As technology continues to transform traditional farming, the integration of logistic regression in precision agriculture offers a practical and efficient approach to crop selection.

Keywords: Machine Learning, Logistic Regression, Crop Prediction, Precision Farming, Agricultural Productivity, Soil and Weather Parameters

1. Introduction

Agriculture, the cornerstone of global food security, faces mounting pressures in the 21st century. A growing population and unpredictable environmental conditions necessitate innovative solutions. Precision farming techniques, harnessing advanced technologies like machine learning, offer a promising path towards optimizing agricultural processes. This research addresses this critical need by developing a machine learning model for crop prediction based on essential soil and weather parameters. Understanding the complex interplay between factors like Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH, and rainfall is fundamental to successful farming. Recognizing the significance of these variables, this study presents a comprehensive analysis employing a logistic regression model. This model will predict the most suitable crops for specific combinations of soil and weather data.

Logistic regression, a powerful classification algorithm, excels at modeling relationships between features and binary outcomes. Here, it will be used to unravel the intricate connections between soil and weather characteristics and their suitability for different crop types. This research delves into the methodology of feature engineering, a crucial process for enhancing the model's predictive capabilities. By optimizing features, we aim to achieve high precision and accuracy in crop recommendations. As we embrace the digital revolution, integrating machine learning into agriculture holds immense potential to transform traditional practices. The outcomes of this research not only offer a practical solution for crop prediction but also pave the way for informed decision-making in resource management and allocation.

2. Related Work

K Nischitha et al.[14] addressed the challenges faced by Indian farmers, who predominantly engage in agriculture without considering optimal crop diversity and fertilizer usage. To mitigate this, they proposed a machine learning-based system that suggests the most suitable crops based on land characteristics and weather parameters. The system further provided precise recommendations for fertilizer content and quantity, facilitating sustainable farming practices. By adopting their system, farmers can diversify crops, potentially increasing profits, while simultaneously reducing soil pollution caused by indiscriminate fertilizer application. This endeavour strived to enhance agricultural practices, fostering a more sustainable and economically viable future for India's farming community.

Racic M et al. [18] investigated the utilization of transfer learning in conjunction with a transformer model for analyzing variable-length satellite image time series (SITS) in crop classification. By integrating remote sensing, machine learning, and agricultural expertise, the aim was to swiftly generate accurate agricultural land maps, reducing the need for interventions and field visits. Sentinel-2 satellite imagery and crop labels spanning from 2019 to 2021 were employed to assess the model's adaptability through fine-tuning for early crop classification. The initial model, trained on data from a different year, achieved an 82.5% F1 score for the target year without specific year references. Transfer learning exhibited notable efficiency, necessitating less than 0.3% of the data for model adaptation, and outperformed the baseline with only 9% of the data after 210 days into the year.

Mohamed-Amine N et al. [10] conducted a focused analysis on phytosanitary treatment sales in Morocco's Souss Massa region, aiming to predict agricultural product sales, specifically crop protection solutions. Data, sourced from an ERP system (Microsoft Dynamics AXAPTA) of a prominent regional agricultural company, included sale dates, farming types, climate, and specific locations. Machine learning, including the Gradient Boosting Regressor algorithm, was employed for forecasting. Model evaluation yielded promising results, with a Mean Absolute Error (MAE) of 0.0035 and Root Mean Square Error (RMSE) of 0.0066. These outcomes, derived from various regression models, contribute valuable insights to agricultural sales prediction, considering climate, farming practices, and regional nuances.

Mohammed S. Arshad et al. [11] tackled the global and regional threat posed by sodium hazard to agricultural production, with a focus on rainwater quality monitoring in central Europe—an aspect previously overlooked. They predicted sodium adsorption ratio (SAR) from 1985 to 2021 using ten ionic species in rainwater and employed four machine learning methods (Random Forest, Gaussian Process Regression, Random Subspace, and Artificial Neural Network-Multilayer Perceptron) at three Hungarian stations. The Mann-Kendall test revealed a significant increasing trend in Na+ ions and SAR, signaling a serious agricultural hazard. The Artificial Neural Network-Multilayer Perceptron emerged as the most effective model, highlighting its potential for improving agricultural water management in Central Europe.

Ramyadevi R et al. [19] addressed the crucial role of insect pest management in enhancing agricultural crop yield and quality. Traditional methods relying on agricultural experts' experience were laborious and subjective. Leveraging the advancements in intelligent techniques, this research employs three cutting-edge Deep Convolutional Neural Network (DCNN) models—Faster-RCNN, Mask-R CNN, and Yolov5—for efficient insect pest detection. Two coco datasets were created based on Baidu AI insect detection and IP102 datasets. Results favor Yolov5 for Baidu AI dataset, with over 99% accuracy and faster computational speed. However, for the complex IP102 dataset, Faster-RCNN and Mask-R CNN outperform Yolov5, achieving a 99% accuracy.

Gowri L et al. [7] explored theoretical approaches, empirical research, and methodological strategies to optimize the nexus's utilization. Addressing water and soil quality degradation, infectious diseases, and human-induced pressures, the book presents an inventory of techniques and practices for climate change adaptation. With a focus on strengths and weaknesses, it served as a crucial resource for scholars, researchers, and decision-makers seeking a deeper understanding of the intricate interactions within the Water-Soil-Plant-Animal nexus and its implications in the face of climate change.

3. Proposed Work

3.1 Machine Learning

Machine learning enables computers to learn from data and make decisions without explicit programming. Its types include supervised learning, unsupervised learning,

reinforcement learning, semi-supervised learning, deep learning, and transfer learning. Understanding these types is essential for leveraging AI effectively.

3.2 Supervised Learning

Supervised learning is a machine learning paradigm where the algorithm is trained on a labelled dataset, consisting of input-output pairs. The goal is for the model to learn the mapping from inputs to corresponding outputs. During training, the algorithm adjusts its parameters to minimize the difference between its predictions and the actual labelled outputs. This enables the model to make accurate predictions on new, unseen data by generalizing patterns learned from the training set. Common applications include classification and regression tasks, making supervised learning a fundamental approach in solving real-world problems.

3.3 Logistic Regression

Logistic regression is a statistical method used for binary classification problems, predicting outcomes that fall into one of two categories. Despite its name, it's employed for classification rather than regression tasks. The algorithm applies a logistic function to linearly combine input features, producing probabilities that an instance belongs to a particular class. Through a process called maximum likelihood estimation, the model optimizes its parameters during training to best fit the observed data. Known for its simplicity, interpretability, and efficiency, logistic regression finds applications in various fields, including medicine, finance, and social sciences, where binary decisions are prevalent. The Figure 1 depicts the components of crop classification.

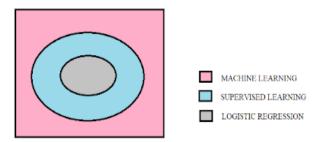


Figure 1. Components of Crop Classification

3.4 Data Collection

Data sources along with the input parameters considered for the prediction of crops is elaborated in this section. To perform the analysis, it is important to gather the datasets from various sources. A dataset of attributes like Nitrogen, Phosphorus, Potassium, Temperature, humidity, pH, rainfall, and crops as labels is collected. It is a dataset of twenty-two different crops and hundred different attribute values for each crop and as a whole 2200 data are there in the dataset. The 22 different crops in the dataset used for prediction are shown in Figure 2.

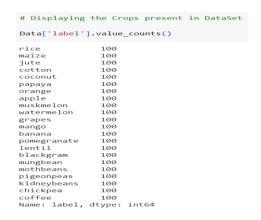


Figure 2. Crops Present in the Dataset

3.5 Preprocessing

The foundation of a robust machine learning model hinges on meticulously prepared data. Pre-processing agricultural data – inherently complex with missing values, outliers, and inconsistencies in scale – becomes crucial. Techniques like imputation address missing entries, ensuring data integrity. Feature scaling levels the playing field for all features by normalizing them to a common range. Outlier detection and handling safeguard the model from the influence of extreme values. Common methods include identifying outliers using interquartile range (IQR) and then removing them, capping their values, or applying winsorization (replacing extremes with values at the distribution's tails).

Feature engineering delves even deeper, crafting informative features that enhance the model's ability to learn. We can create new features, like a soil health index based on NPK values, or transform existing ones through techniques like log transformations to normalize skewed distributions. Categorical variables, like crop types, are encoded into numerical formats

using one-hot encoding. Finally, splitting the pre-processed data into training and testing sets is paramount. The training set guides the model's learning, while the testing set provides an unbiased evaluation of its generalizability on unseen data. By diligently applying these techniques, we cultivate a high-quality dataset that empowers the machine learning model to generate accurate and insightful crop predictions. Figure 3 illustrates the flow graph of the proposed model.

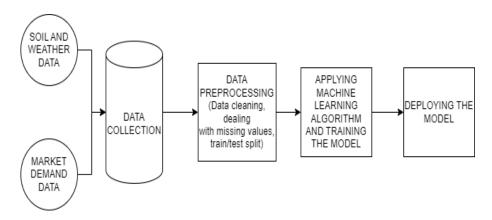


Figure 3. Flow Graph of the Proposed Model

3.6 Architecture of Logistic Regression

Logistic regression's architecture is instrumental in predicting the suitability of a particular crop based on various factors. The input layer represents essential agricultural features such as soil composition, temperature, humidity, and rainfall. Each of these features is assigned a weight during the model training process, reflecting their significance in determining the optimal crop choice. The weighted sum, similar to the linear combination in logistic regression, captures the relationships between these features and the likelihood of a specific crop being suitable.

The activation function, typically a sigmoid in logistic regression, transforms this linear combination into a probability score. This probability score could be interpreted as the likelihood of a particular crop being well-suited for the given set of agricultural conditions. The final output layer then classifies the crop based on a chosen threshold. If the probability exceeds a certain threshold, the model suggest that the conditions are conducive for cultivating a specific crop, and vice versa.

During the training phase, the logistic regression model adjusts its weights to optimize predictions. This adaptation process ensures that the model becomes adept at recognizing patterns in the input features and making accurate predictions regarding suitable crops. In essence, the logistic regression architecture serves as a foundational framework for developing a model to suggest a crop, leveraging its simplicity and interpretability to provide valuable insights for farmers in optimizing their agricultural practices.

3.7 Training

Training a dataset for developing a logistic regression model involves a series of crucial steps. Initially, the dataset is split into training and testing sets, with the former serving as the foundation for model learning by allocating 80% of the data to the training set and the remaining 20% to the testing set. Feature scaling is then applied to normalize numeric attributes, preventing any single feature from dominating the training process. Addressing missing values through techniques like imputation ensures data integrity. Categorical variables are encoded into numeric format, typically using one-hot encoding. The core of the process lies in model training, where logistic regression adjusts its weights to minimize the disparity between predicted probabilities and actual class labels.

3.8 Testing and Model Evaluation

In logistic regression, testing and assessment are essential processes to guarantee the model's dependability and its capacity to generalize to new data. Using a different testing dataset, the model's performance is evaluated following training. Metrics such as recall, accuracy, precision, and F1 score are commonly employed to assess the Logistic regression model's performance in predicting suitable crops on both the trained dataset and unseen data.

3.9 Tools Used

Programming Language: Python is chosen for machine learning due to its extensive libraries and ease of use.

Machine Learning Libraries: Libraries like scikit-learn (Python) for implementing machine learning algorithms like logistic regression.

Data Analysis Libraries: Libraries like pandas (Python) are used for data manipulation and pre-processing tasks.

Data Visualization Libraries: Libraries like matplotlib (Python) used to create informative visualizations of data and model results.

4. Results and Discussion

The accuracy of the developed model is 96.8%. As the model performs a multiclass classification, the precision, recall and F1-score are calculated by taking average as micro, macro and weighted as the values. 'Macro' computes the value for each class independently and then takes the average. 'micro' computes the value globally by counting the total true positives, false negatives, and false positives. 'weighted' is similar to 'macro' but accounts for class imbalance. The value of accuracy of the model is 96.8%. The value of precision micro is 96.8%, precision macro is 97%, precision weighted is 96.9%. The value of recall micro is 96.8%, recall macro is 97%, recall weighted is 96.8%. The value of F1-score micro is 96.8%, F1-score for macro is 96.9%, F1-score for weighted is 96.8%. The confusion matrix of the proposed Logistic regression architecture is given in Figure 4.

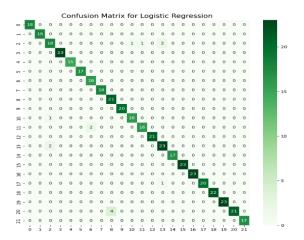


Figure 4. Confusion Matrix of the Logistic Regression Model

The output statistics details of the selected crop (cotton) displayed is shown in Figure 5.

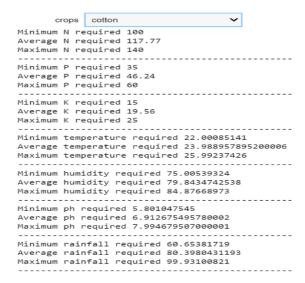


Figure 5. Output for the Selected Crop (cotton)

The crops that can be cultivated in every particular condition are given in Figure 6(a) to Figure 6(g).



Figure 6. (a) Output for the Condition 'Temperature'





Figure 6. (B) Output for the Condition 'Rainfall' Figure 6. (c) Output for the Condition 'N'

conditions P

Average value for P is 53.36

Rice: 47.58

Rice: 47.58

Riack Grams: 67.47

Banana: 82.01

Jute: 46.86

Coconut: 16.93

Apple: 134.22

Papaya: 59.05

Muskmelon: 17.72

Grapes: 132.53

Watermelon: 17.00

Kidney Beans: 67.54

Mung Beans: 47.28

Oranges: 16.55

Chick Peas: 67.79

Lentils: 68.36

Cotton: 46.24

Maize: 48.44

Moth Beans: 48.01

Pigeon Peas: 67.73

Mango: 27.18

Pomegranate: 18.75

Coffee: 28.74



Figure 6. (d) Output for the Condition 'P' Figure 6. (e) Output for the Condition 'Potassium'

conditions ph

Average value for ph is 6.47

Rice: 6.43
Black Grams: 7.13
Banana: 5.98
Jute: 6.73
Coconut: 5.98
Apple: 5.93
Papaya: 6.74
Muskmelon: 6.36
Grapes: 6.03
Watermelon: 6.50
Kidney Beans: 5.75
Mung Beans: 6.72
Oranges: 7.02
Chick Peas: 7.34
Lentils: 6.93
Cotton: 6.91
Maize: 6.25
Moth Beans: 6.83
Pigeon Peas: 5.79
Mango: 5.77
Pomegranate: 6.43
Coffee: 6.79



Figure 6. (f) Output for the Condition 'pH' **Figure 6.** (g) Output for the Condition 'Humidity'

4.1 Benefits and Applications

(a) Enhanced Resource Management

The project facilitates optimized resource allocation for farmers by providing accurate predictions on suitable crops based on soil and weather conditions. This empowers farmers to strategically allocate resources such as fertilizers and irrigation, minimizing waste and maximizing efficiency in agricultural practices. As climate patterns change, the logistic regression model assists farmers in adapting their crop choices to new conditions. This proactive approach promotes resilience in the face of climate change, allowing farmers to adjust their practices accordingly.

(b) Increased Agricultural Productivity

Accurate crop predictions contribute to higher crop yields. Farmers can make informed decisions about crop selection, ensuring that chosen crops are well-suited to the prevailing environmental conditions. This, in turn, leads to increased agricultural productivity. The project serves as a risk mitigation tool for farmers by providing insights that reduce the risk of crop failure. By selecting crops that are resilient to specific environmental factors, farmers can mitigate potential losses and enhance the overall resilience of their agricultural practices.

(c)Sustainable Agricultural Practices

Aligning crop choices with environmental conditions promotes sustainable agricultural practices. The project encourages farmers to make decisions that minimize the environmental impact of farming activities, contributing to long-term sustainability. The predictive nature of the model saves farmers time and costs associated with trial-and-error approaches. By leveraging data-driven insights, farmers can make efficient and economical decisions in their farming practices.

(d) Supporting Precision Farming

The application of the logistic regression model supports the principles of precision farming. By tailoring agricultural practices to specific soil and weather conditions, farmers can achieve optimal results and minimize resource wastage. The model's suggestions for suitable

crops enhance crop rotation strategies. This is crucial for maintaining soil health, preventing diseases, and optimizing the long-term sustainability of agricultural land.

5. Conclusion

The potential for changing agricultural methods is enormous when using logistic regression to predict viable crops based on weather and soil conditions. This study emphasizes how important it is to use machine learning algorithms in contemporary agriculture in order to maximize crop production and minimize waste by making well-informed decisions. The opportunities and difficulties of implementing these technologies in agriculture are identified in this study. The results imply that farmers may be able to make better judgments on the variables influencing crop growth by examining a broad range of data gathered from farms. The results highlight the real advantages it offers farmers, from improved agricultural output and resource management to risk reduction and the encouragement of sustainable farming. The accuracy of crop predictions not only empowers farmers with informed decision-making but also contributes to the broader goals of precision farming and environmental sustainability. As we witness the impacts of climate change on agriculture, the model's capability to suggest crops resilient to evolving conditions becomes paramount for ensuring long-term sustainability.

References

- [1] Agrawal, N., Govil, H., & Kumar, T. (2024). Agricultural land suitability classification and crop suggestion using machine learning and spatial multicriteria decision analysis in semi-arid ecosystems. Environment, Development and Sustainability, 1-38.
- [2] Anguraj, K., Thiyaneswaran, B., Megashree, G., Shri, J. P., Navya, S., & Jayanthi, J. (2021). Crop recommendation on analyzing soil using machine learning. Turkish Journal of Computer and Mathematics Education, 12(6), 1784-1791.
- [3] Akulwar, P. (2020). A recommended system for crop disease detection and yield prediction using machine learning approach. Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural and Other Industries, 141-163.

- [4] Banerjee, S., & Mondal, A. C. (2023). A Region-Wise Weather Data-Based Crop Recommendation System Using Different Machine Learning Algorithms. International Journal of Intelligent Systems and Applications in Engineering, 11(3), 283-297.
- [5] Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., ... & Saker, L. (2023). Crop prediction model using machine learning algorithms. Applied Sciences, 13(16), 9288.
- [6] Gosai, D., Raval, C., Nayak, R., Jayswal, H., & Patel, A. (2021). Crop recommendation system using machine learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 7(3), 558-569.
- [7] Gowri, L., Pradeepa, S., & Sasikaladevi, N. (2024). Weather-Based Crop Prediction Using Big Data Analytics. In Water-Soil-Plant-Animal Nexus in the Era of Climate Change (pp. 78-92). IGI Global.
- [8] Kalimuthu, M., Vaishnavi, P., & Kishore, M. (2020, August). Crop prediction using machine learning. In 2020 third international conference on smart systems and inventive technology (ICSSIT) (pp. 926-932). IEEE.
- [9] Mishra, S., Mishra, D., & Santra, G. H. (2016). Applications of machine learning techniques in agricultural crop production: a review paper. Indian J. Sci. Technol, 9(38), 1-14.
- [10] Mohamed-Amine, N., Abdellatif, M., & Belaid, B. (2024). Artificial intelligence for forecasting sales of agricultural products: A case study of a moroccan agricultural company. Journal of Open Innovation: Technology, Market, and Complexity, 10(1), 100189.
- [11] Mohammed, S., Arshad, S., Bashir, B., Vad, A., Alsalman, A., & Harsányi, E. (2024).
 Machine learning driven forecasts of agricultural water quality from rainfall ionic characteristics in Central Europe. Agricultural Water Management, 293, 108690.
- [12] Motwani, Aditya, Param Patil, Vatsa Nagaria, Shobhit Verma, and Sunil Ghane. "Soil Analysis and Crop Recommendation using Machine Learning." In 2022 International Conference for Advancement in Technology (ICONAT), pp. 1-7. IEEE, 2022.

- [13] Musanase, C., Vodacek, A., Hanyurwimfura, D., Uwitonze, A., & Kabandana, I. (2023). Data-Driven Analysis and Machine Learning-Based Crop and Fertilizer Recommendation System for Revolutionizing Farming Practices. Agriculture, 13(11), 2141.
- [14] Nischitha, K., Vishwakarma, D., Ashwini, M. N., & Manjuraju, M. R. (2020). Crop prediction using machine learning approaches. International Journal of Engineering Research & Technology (IJERT), 9(08), 23-26.
- [15] Pathak, P., Phalke, R., Patil, P., Prakash, P., & Bhosale, S. (2020). Survey on Crop Suggestion Using Weather Analysis. International Research Journal of Engineering and Technology (IRJET). 7(01), 1045-1047
- [16] Prakash, R. V., Abrith, M. M., & Pandiyarajan, S. (2022, May). Machine learning based crop suggestion system. In 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1355-1359). IEEE.
- [17] Priya, P. K., & Yuvaraj, N. (2019). A survey on deep learning based IoT approach for precision crop suggestion. International Journal for Research in Applied Science & Engineering Technology, 7(2), 389-395.
- [18] Racic M., Oštir, K., Zupanc, A., & Čehovin Zajc, L. (2024). Multi-Year Time Series Transfer Learning: Application of Early Crop Classification. Remote Sensing, 16(2), 270.
- [19] Ramyadevi, R., Sandeep, K., Sairam, V. G., & Vaisak, P. (2024). A system for crop recommendation to improve the production management. In Artificial Intelligence, Blockchain, Computing and Security Volume 2 (pp. 213-217). CRC Press.
- [20] Saranya N., & Mythili, A. (2020). Classification of soil and crop suggestion using machine learning techniques. International Journal of Engineering Research & Technology (IJERT), 9(02), 671-673.