



RESEARCH ARTICLE

MACHINE LEARNING BASED CROP, YIELD AND PRICE PREDICTION SYSTEM FOR SMART AGRICULTURE USING RANDOM FOREST AND WEATHER PARAMETERS

Dr. Harish S Gujjar

Associate Professor and Head, Department of Computer Science, SSAS GFGC Hosapete – 583201, Karnataka, India

ARTICLE INFO

Article History:

Received 18th February, 2026
Received in revised form
24th March, 2026
Accepted 20th April, 2026
Published online 29th May, 2026

Key Words:

Machine Learning, Crop, Yield, Price Prediction, Smart Agriculture.

*Corresponding author:

Dr. Harish S Gujjar

ABSTRACT

This research paper focuses on reducing agricultural losses caused by changing climatic and environmental conditions, which often result in unsuitable crop selection for cultivation on a particular land area. Machine Learning ML is a significant methodology for accomplishing the reasonable and compelling answers for this disadvantage. In India farmers still follow the traditional technology which they adopted from their ancestors. But the problem is that in the earliest time climate was very healthy everything happened on time. But now most of the things have been changed due to global warming and many other factors. The leading annoyance with agribusiness in India is the shortage of rainfall in seasonal periods. Humidity is also required for production, though it has been unreasonable, it also transforms as a weakness. Accurateness of harvest price forecasting strategies plays an important function in encouraging market characteristics such as direction and collection. Harvest Yield Prediction technique includes foreseeing yield of the harvest from reachable historical and possible data like climate parameter, soil parameter and yield prediction. Machine learning algorithms are applied and we get best predicted results through a web application. This Project is an attempt of predicting the outcome of harvest supported the current data by using of RFA Random Forest Algorithm and Back Propagation. The expectation can make the farmer to foresee the yield of harvest before developing onto the agribusiness zone.

Copyright©2026, Harish S Gujjar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Dr. Harish S Gujjar. 2026. "Machine Learning Based Crop, Yield and Price Prediction System for Smart Agriculture using Random Forest and Weather Parameters". *International Journal of Current Research*, 16, (05), 37167-37170.

INTRODUCTION

Machine Learning (ML), a subset of Artificial Intelligence (AI), refers to the ability of computer systems to learn from experience and improve their performance without being explicitly programmed. ML systems are trained on historical data—referred to as *training data*—to identify patterns or make predictions. These trained models are then evaluated on unseen data, known as *testing data*, to assess their generalization capability. Broadly, ML techniques are classified into three categories: Supervised Learning, where models are trained using labeled data; Unsupervised Learning, which involves discovering hidden patterns in unlabeled data through clustering or dimensionality reduction; and Reinforcement Learning, where agents learn optimal strategies through interaction with an environment based on feedback or rewards. ML algorithms have found widespread applications across domains such as image recognition, email filtering, fraud detection, and computer vision. In this study, ML techniques are employed to classify animal species from images based on their features, enabling automated detection and classification for ecological and agricultural applications.

Python has become the language of choice in ML research and application due to its simplicity, readability, and a vast ecosystem of scientific libraries. Its support for platform independence enables seamless code portability across systems including Windows, mac OS, Linux, and Unix. Python offers access to powerful libraries such as NumPy, pandas, scikit-learn, and Matplotlib, which streamline data preprocessing, statistical analysis, and visualization. For deep learning and large-scale model development, libraries like TensorFlow and PyTorch are commonly used, allowing researchers to build and deploy complex models efficiently. Among the tools used in this project are:

- **Open CV:** An open-source computer vision library that supports real-time image processing tasks such as object detection, face recognition, and tracking. It offers interfaces in C++, Python, and Java.
- **Tensor Flow:** A robust, end-to-end machine learning platform developed by Google, known for its flexibility and scalability. TensorFlow supports the creation, training, and deployment of ML models and is widely used in both research and production environments.

- **Ensembl:** Though primarily a bioinformatics platform, Ensembl's systematic data structuring serves as an inspiration for handling complex biological datasets, which parallels the way agricultural datasets can be curated for prediction tasks.

Agriculture remains a vital sector in India, contributing approximately 18% to the national Gross Domestic Product (GDP) and employing over 50% of the workforce. Despite its economic significance, the agricultural sector has experienced a declining share in national output, underscoring the need for technological interventions to improve productivity and sustainability. Accurate forecasting of crop yield and pricing is essential for policymakers, stakeholders, and farmers to make informed decisions regarding production, supply chain management, and market strategies. This research aims to leverage machine learning techniques to predict crop yields and prices across various regions in India. By integrating historical agricultural data, climatic variables, and economic indicators, the study constructs predictive models that assist in estimating harvest outcomes. A novel weather index and technological factors are also employed as input features. Model evaluation is carried out using performance metrics such as Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Average Error Rate, enabling comparison between actual outcomes (targets) and predicted values. Ultimately, the goal of this research is to build a reliable and accessible prediction framework that can guide farmers, improve agricultural planning, and contribute to sustainable development in India's agrarian economy.

Literature Survey

Crop price and yield prediction has garnered significant attention in recent years due to its direct impact on agricultural planning, economic sustainability, and food security. Numerous studies have explored various machine learning (ML) and time-series forecasting techniques to address the challenge of accurate and timely agricultural prediction. In a study by Singh *et al.* (2021), a Long Short-Term Memory (LSTM)-based hybrid model was developed to predict crop prices using parameters such as crop type, sowing date, area harvested, and historical price data. The research utilized datasets from *data.gov.in* and *agmarknet.gov.in*, specifically focusing on the Pune region. After merging time-series data based on crop name, year, and month, correlation analysis was performed, and low-correlation features were removed. Algorithms such as Random Forest Regressor, KNN, Decision Tree Regressor, and MLP were evaluated using the LazyPredict library (1). Kumar *et al.* (2021) proposed a forecasting system that employed decision trees and other supervised learning models to predict future crop prices based on variables such as rainfall, minimum support price (MSP), and farming costs. The emphasis was on supporting farmers and agricultural industries through early price forecasts, thereby maximizing profits and improving planning efficiency (2). Joseph *et al.* (2020) focused on areca nut price forecasting in Kerala using SARIMA, Holt-Winters, and LSTM models. The study, based on data from 2007–2017, demonstrated that LSTM performed best in terms of RMSE. Data preprocessing included handling missing values and identifying strong predictor variables using a correlogram with a threshold correlation of 0.7 (3). An economic forecasting framework was introduced by Desai *et al.* (2020) to address inconsistencies in market-recorded crop prices caused by human error and

volatility. This model integrated historical prices, climatic data, and transportation conditions, emphasizing robust ML-based strategies for reliable price predictions (4). Patel *et al.* (2020) designed a machine learning-based system using XGBoost, Linear Regression, and Neural Networks to predict crop prices with high accuracy. The study concluded that XGBoost provided optimal results, and RMSE was used as a key metric to evaluate model performance (5). Gupta *et al.* (2021) developed a web-based decision support tool using a Decision Tree Regressor, aimed at forecasting crop prices for over 20 crops. The system integrated rainfall and monthly market data to deliver actionable insights through an interactive dashboard (6). In a pilot study conducted in Madhya Pradesh, Sharma *et al.* (2020) avoided crops like wheat and paddy due to their price regulation by MSP. Their ML pipeline involved data preparation, training multiple algorithms, and conducting live testing to refine model accuracy for volatile crops (7). An advanced approach to model selection under volatile conditions was proposed by Mehta *et al.* (2022), who employed feature concatenation (FC) and time-series-dependent (TD) features. Their model, when tested on tomato and maize, showed improved predictive performance by accounting for market arrivals and data quality variations (8). Reddy *et al.* (2020) proposed a decision-tree-based model trained on Kharif crops such as paddy and barley. The tool offered both yield and price predictions and advocated the integration of satellite imagery and AI to support real-time agricultural monitoring (9). A flash-based ML platform by Banerjee *et al.* (2021) utilized Decision Tree Regressor models to predict prices of major Kharif and Rabi crops. The model accounted for climatic variations, fertilizer recommendations, and disease alerts, contributing to smart agriculture (10).

Karthik *et al.* (2020) employed Random Forests on time-series crop summary datasets for both global and regional yield predictions. Data was split into training and testing sets (75:25), and supervised learning methods were applied for accurate label prediction (11). Rao *et al.* (2021) focused on yield prediction using a Random Forest classifier. The model facilitated optimal crop selection and strategic agricultural planning by analyzing one-time crop data, with emphasis on integrating soil information and irrigation parameters (12). An Interaction Regression Model presented by Liu *et al.* (2021) demonstrated superior accuracy in predicting soybean and grain yields across U.S. Midwest states. It uniquely highlighted the importance of climate-soil-management interactions and the potential for incorporating genetic and satellite data (13). In Malaysia, Sulaiman *et al.* (2021) assessed the impact of El Niño on oil palm yields by introducing the Fresh Fruit Bunch Index (FFBI), which outperformed traditional models in predictive accuracy and demonstrated significant economic implications (14). Toth *et al.* (2021) analyzed maize hybrid performance in Hungary under different fertilizer regimes. The study emphasized sustainable practices and revealed that nitrogen fertilizer had the most substantial impact on yield stability (15). A related study by Nagy *et al.* (2021) examined the effect of NP fertilizer placement depth on maize emergence using Additive Main Effects and Multiplicative Interaction (AMMI) analysis. They concluded that a depth of 10 cm was optimal for plant emergence stability (16). Chatterjee *et al.* (2021) applied both Multiple Linear Regression and Artificial Neural Networks to predict potato yields before harvest. Neural networks outperformed regression models with a Mean Absolute

Percentage Error (MAPE) of 7.2%, enabling more informed pre-harvest decision-making (17). A large-scale Indian study by Bansal *et al.* (2020) utilized stacked regression techniques, including Lasso, Elastic Net, and Kernel Ridge, for crop yield prediction across over 2.5 lakh observations. The stacked approach achieved sub-1% RMSE error (18). Kumar and Singh (2020) applied ML algorithms such as Decision Trees, Lasso, and Ridge Regression to estimate yields based on location, historical yield, and soil nutrients. Their findings affirmed the superiority of Decision Trees in prediction accuracy (19). Finally, Ahmed *et al.* (2020) implemented Support Vector Regression and Linear Regression on 140 data points to forecast yield. The study reported an MSE of 0.005 and an R^2 of 0.85, indicating reliable performance across varied soil and climate conditions (20).

Problem statement: The production of agriculture is affected by several climate factors. Like as crop price, metrological parameters, precipitation parameters, and soil parameters. And due to continuous change in climate conditions, everything is messed. In India agriculturalists still follow the classic technology which they assumed from their predecessors. But the problem is that before the time the weather was very healthy everything occurred on time. But now most of the things have been changed due to global warming and many other factors and market demand. The detailed and accurate information about the crop and price across various states helps us to transform our ability to manage and implement. Here we investigate the crop and yield data collected by various states and average price of the crop and its location using machine learning technique.

Objective of the paper: Advancements in Artificial Intelligence and implementing the way people use the technology with crop and price prediction in agriculture the learning of the connotations of the concepts, access to the latest approaches, methods, and theory. Realizing the investigation topics based on the current research Concentrate on your domain of expertise even if one field uses the same words, they usually represent completely. And furthermore enhances the differentia of the publication's survey to restrict sidetracks and recognize to illustrate what is excluded.

- Prediction of crop price based on the year, season, crop, state, district parameter with comparing the previous datasets value.
- Prediction of crop based on pH value, NPK values, rainfall, state and district parameters from the datasets.
- Prediction of crop-yield based on the year, season, crop, temperature, humidity, soil- moisture from the datasets.
- Parameters changes based on the objective of the user.

METHODOLOGY

A machine learning system is a computer program that learns and improves based on experience without needing to be explicitly programmed. A machine learning program uses data to learn for itself by accessing it and using it. As we learn, we look for patterns in our observations, our direct experience, or instructions in order to make better decisions in the future. This process begins with observations or data, such as examples, direct experience, or instruction. In order for the computer to learn automatically and adjust actions accordingly, it must be able to learn without human assistance.

Dataset Description: Generally researchers used .csv files of agriculture dataset for crop yield prediction. The dataset is supervised learning. It consists of different attributes like State, district, humidity, temperature, NDVI, wind Speed, yield etc.

Architecture Description: Architecture is the conceptual model that describes how a system works, what it looks like, and what it is capable of. A system architecture description describes the structure and behaviour of the system in a way that facilitates analysis and reasoning. System architecture consists of the components and subsystems that will play a role in the implementation of the overall system. There have been improvements in the field of Architecture Description Languages. Architecture description languages are formal languages which are used to describe the system architecture. In the figure below, you can see that our working model contains different modules connected by Python programming language which is as shown in figure 1.

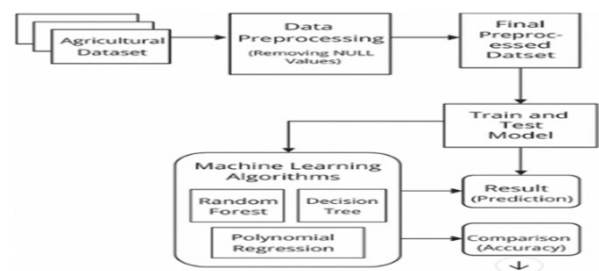


Fig. 1. System Methodology

The proposed system is designed with five integrated modules that support the prediction of crops, crop yield, and market prices through Machine Learning methods. Important techniques applied in the system include data cleaning and preprocessing, graphical data analysis, Multiple Linear Regression, Random Forest methods, dataset training and evaluation, and accuracy measurement using Mean Absolute Percentage Error (MAPE). The framework also incorporates IoT-based sensor inputs, cloud connectivity, and weather information obtained through APIs to improve prediction performance. Separate modules are available for crop recommendation, yield estimation, and price forecasting, offering an easy-to-use interface for users. The complete implementation is developed using Python along with tools and libraries such as Google Colab, OpenCV, and NumPy for efficient computation and data handling.

CONCLUSION

We have proposed an approach of Predicting the crop, crop-yield and crop-price using RFA Random Forest and Back Propagation Algorithm. To improve Agribusiness, this approach reduces farmers' losses and improves yields. In this phase of the project, we reviewed the literature on the price prediction, crop-yield prediction. This literature helps us understand the challenges that we face in the price dataset to identify the price and yield. According to the literature experimental results we came to know that random forest and back propagation helps in crop-price Identification and helps to eradicate the challenges that we have mentioned in the dataset that comes in the dataset and also the accuracy rate of this Algorithm is far better than other Algorithm. By integrating this with other departments like sericulture and other development initiatives on the village level, farmers are

able to gain a deeper understanding of agricultural ecosystems. By incorporating this with other departments, such as sericulture, this can be improved in order to develop villages for the benefit of farmers.

REFERENCES

1. Kiran M. Sabu, T.K. Manoj Kunmar, Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala, *Procedia Computer Science*.
2. Applied different Algorithm comparison algorithms-you-should-know-introductory-guide
3. Haque, F. F. A. Abdelgawad, V. P. Yanambaka and K. Yelamarthi, "Crop Yield Analysis Using Machine Learning Algorithms," 2020 IEEE 6th World Forum on Internet of Things (WF-IoT), 2020, pp. 1-2, doi: 10.1109/WFloT48130.2020.9221459.
4. Kavita M. and P. Mathur, "Crop Yield Estimation in India Using Machine Learning" 2020 IEEE Sth International Conference on Computing Communication and Automation (ICCCA), 2020, pp. 220-224, doi: 10.1109/ICCCA49541.2020.9250915
5. Nishant, P. S. P. Sai Venkat, B. L. Avinash and B. Jabber, "Crop Yield Prediction based on Indian Agriculture using Machine Learning," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-4, doi: 10.1109/INCET49848.2020.9154036.
6. Kumar, R. M. P. Singh, P. Kumar and J. P. Singh, "Crop Selection Method to maximize crop yield rate using machine learning technique," 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (1CSTM), 2015, pp. 138-145, doi: 10.1109/1CSTM.2015.7225403.
7. Analysis Of Crop Yield Prediction Using Data Mining Techniques (2015).
8. Effects of NP Fertilizer Placement Depth by Year Interaction on the Numeral of Maize (*Zea mays* L.) Plants after Emergence Using the Additive Main Effects and Multiplicative Interaction Model (July 2021).
9. Bojtor, C.; Mousavi, S.M.N.; Illés, A.; Széles, A.; Nagy, J.; Marton, C.L. Stability and Adaptability of Maize Hybrids for Precision Crop Production in a Long-Term Field Experiment in Hungary. *Agronomy* 2021, 11, 2167. <https://doi.org/10.3390/agronomy1112167>
10. Khor, Jen Feng & Ling, Lloyd & Yusop, Zulkifli & Tan, Weilun & Ling, Joan & Soo, Eugene. (2021). Impact of El Niño on Oil Palm Yield in Malaysia. *Agronomy*. 11. 2189. [10.3390/agronomy11112189](https://doi.org/10.3390/agronomy11112189). (2021).
11. Ansarifar, J., Wang, L. & Archontoulis, S.V. An interaction regression model for crop yield prediction. *Sci Rep* 11, 17754 (2021). <https://doi.org/10.1038/s41598-021-97221-7> (12). Crop Yield Prediction using Machine Learning Algorithms (2021)
12. Ranjani Dhanapal *et al* 2021 *J. Phys.: Conf. Ser.* 1916 012042 Crop price prediction using supervised machine learning algorithms.
13. Jain, Ayush & Marvaniya, Smit & Godbole, Shantanu & Munigala, Vitobha. (2022). Towards Context-based Model Selection for Improved Crop Price Forecasting. [10.1145/3493700.3493725](https://doi.org/10.1145/3493700.3493725). (January 2022).
14. AIGGPA; Crop price prediction using machine learning in Madhya Pradesh; A pilot study, Atal Bihari Vajpayee Institute of Good Governance and Policy Analysis, March 2020.
15. Ghutake, I, Verma, R.K., Chaudhari, R., & Amarsinh, V. (2021). An intelligent Crop Price Prediction using suitable Machine Learning Algorithm. *ITM Web of Conferences*.
16. Ashoka, D. V., Abhilash, C. B., & Bharath, M. B. (2019, March). A Framework for Enhancing the Agriculture yield using Cloud Clusters. In 2019 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE) (pp. 162-166). IEEE.
17. Chetan, R., Ashoka, D.V., Ajay Prakash, B.V. (2021). Smart Agro-Ecological Zoning for Crop Suggestion and Prediction Using Machine Learning: An Comprehensive Review. In: Chiplunkar, N., Fukao, T. (eds) *Advances in Artificial Intelligence and Data Engineering. Advances in Intelligent Systems and Computing*, vol 1133. Springer, Singapore. https://doi.org/10.1007/978-981-15-3514-7_94.
