International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 3, 2025
https://theaspd.com/index.php

Artificial Intelligence Driven Crop Price and Yield Prediction Using ML for Sustainable Agriculture

Bhavani K G[1], Dr. Vani N[2]
[1]*PG Scholar, Department of Computer Science and Engineering, Adichunchanagiri University, BGS Institute Of Technology, BG Nagar, Karnataka, India*
[2]*Associate Professor, Department of Computer Science and Engineering, Adichunchanagiri University, BGS Institute Of Technology, BG Nagar, Karnataka, India*

This work suggests a machine learning based approach to improve agricultural decision-making by utilising real-time data to increase yield accuracy, lower climate-related risks, and support sustainable farming practices, which uses Polynomial Regression (PR) and Random Forest Algorithm (RFA) to forecast crop production in India. Materials and methods. In order to examine crop production, crop recommendation and price prediction patterns in India, the materials and methodologies employed in the dataset from the given Crop prediction URL, crop recommendation URL and fertilizer prediction URL include the use of multivariate datasets that comprise numerical data and categorical data. In order to train machine learning models, the dataset is pre-processed by handling missing values, encoding categorical variables, and normalising or scaling numerical features. Based on these many characteristics, models such as RFA and Machine Learning (ML) are commonly employed to forecast crop yield.

Results and discussion. *Model performance is evaluated by comparison of predicted values with time real results. Metrics such as MAPE and RMSE are used to assess accuracy. The best-performing model is selected for implementation based on these evaluations. In this study, projected values and real-time outcomes were compared to assess the model's performance. Accuracy was evaluated using metrics including Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). With a MAPE of 8.12% and an RMSE of 24.5, the RFA outperformed the Decision Tree (MAPE: 10.45%, RMSE: 30.2) and PR (MAPE: 9.25%, RMSE: 27.8). These assessments demonstrated the Random Forest model's greater prediction accuracy, which led to its selection for deployment.*

Conclusion. *Our approach helps farmers make better decisions, lower financial risks, and increase agricultural output by utilizing both historical and current data. We found major obstacles to pricing and yield prediction, including climate variability and inconsistent data, by conducting a thorough analysis of the body of current work. Farmers may get a greater understanding of market trends, soil health, and climatic patterns by integrating AI-driven agricultural forecasts with village-level development initiatives.*

Keywords:*Climate variability, Crop selection, /Yield prediction, Random Forest Algorithm (RFA), Back Propagation, Data-driven decision-making, Agricultural productivity, Risk mitigation, Modern farming techniques*

INTRODUCTION

A key pillar of the Indian economy, the agriculture industry employs around half of the country's workers and accounts for 18% of its GDP. But according to new research, agricultural output is gradually declining, in part because of unstable markets, antiquated farming methods, and environmental concerns. Even though farming is one of India's most important economic sectors, most farming still uses antiquated practices that are typically out of step with contemporary demands. For agricultural decision-making to be efficient, precise, real-time crop needs monitoring and production projections are essential. This paper provides a machine learning (ML)-based crop and price prediction model that uses Multiple Linear Regression (MLR) and the Random Forest Algorithm (RFA) to improve forecast accuracy. To provide accurate output and price forecasts, the algorithm examines past crop yield data, soil properties, climate, and market trends.

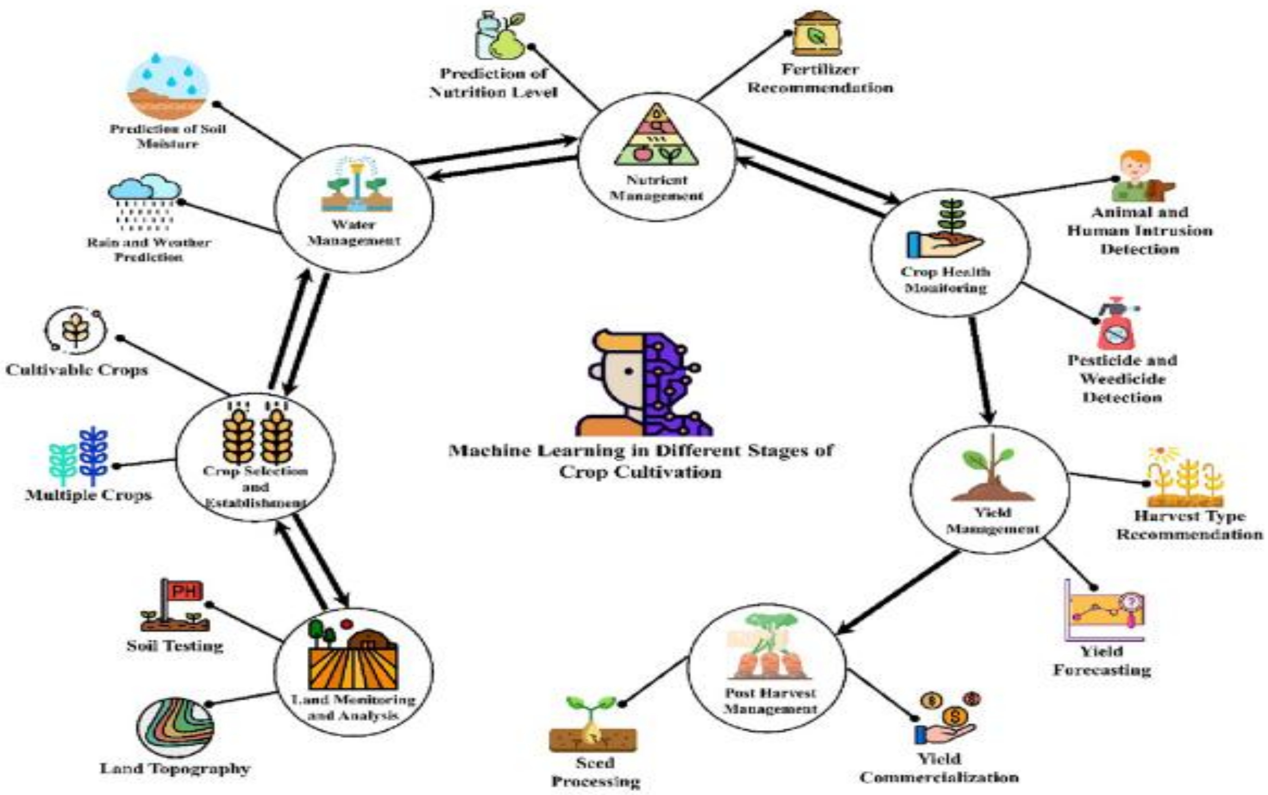


Figure 1 Applications of ML in Agriculture

Regular weather patterns used to allow for consistent farming cycles, but environmental instability and global warming have broken this balance. As a result, farmers deal with resource abuse, financial losses, and unreliable agricultural yields. A data-driven approach is needed to tackle these challenges, enabling farmers to make decisions based on the most recent data on the climate, soil, and economics. Combining machine learning techniques with agricultural yield forecasting is the aim of this study. Through the analysis of multi-dimensional agricultural data, the system generates predictive insights that improve farming operations and reduce climate-induced hazards.Overall working of AI-Driven Crop Prediction and Price Forecasting System is shown in Figure 1. A machine learning approach was presented in [1] to forecast the best time to plant crops by examining meteorological data. For precise predictions, the model made use of the RF, Decision Tree, and Polynomial Regression methods. Additionally, researchers examined the application of ML methods for yield forecast and nitrogen-level estimate in [2], highlighting the significance of data driven insights in enhancing agricultural output. The need of real-time monitoring for sustainable farming was highlighted by Dimitriadis et al. [3], who presented a machine learning founded crop organization model that examined plant health and water requirements. An analysis of how environmental changes affect agricultural productivity was the subject of another research [4]. Crop productivity and selection are influenced by a number of variables, such as weather patterns, soil health, geographic location, ambient conditions, and nutrient content (e.g., pH value, NPK levels). Nevertheless, the majority of current models only include weather or soil characteristics separately. Developing a sophisticated machine learning-based crop selection model that incorporates soil and meteorological characteristics for accurate crop recommendations and optimal yield optimization is the main goal of this project. AI-Powered Prediction: Using both historical and current data, use machine learning algorithms to predict agricultural output and market pricing.Smart Crop Selection: Use environmental, climatic, and soil factors to suggest the best crops for increased yield.IoT-Driven Decision Support: Combine cloud computing with IoT sensors to provide automated agricultural insights and real-time monitoring.

MATERIALS AND METHODS

Machine Learning-Based Crop Price Prediction

A method for predicting crop prices using LSTM models was presented by the LSTM founded Crop Price Forecast System (2021).[5] The study examined variables including location and planting date with an emphasis on unusual crops. Hybrid sequential modeling and conventional machine learning models were the two methods that were assessed. After filtering pertinent factors including crop year, area, and yield, the dataset merged time-series data from data.gov.in and agmarknet.gov.in. Using Lazy Predict for automated model fitting, several regression methods were tried, such as RF and DT.

In order to reduce forecasting mistakes, the Cluster Prediction and Forecasting System utilizing Supervised ML (2021) investigated Decision Tree regression. To guarantee precise price forecasts, factors including rainfall, minimum support prices (MSP), and farming expenses were taken into account.[6] The system's early crop price estimates were intended to help farmers with financial management and strategic planning. Using time-series forecasting models like SARIMA, Holt-Winters, and LSTM neural networks, the 2020 study Estimations in Agriculture: Foreseeable Prices for Arecanuts in Kerala examined monthly price patterns for arecanuts. The Department of Economics and Statistics of Kerala provided the dataset, which covered the years 2007–2017.[7] To increase accuracy, data preparation methods were used, such as interpolation for missing values. In terms of predictive performance, the LSTM model fared better than alternative methods. Price forecast accuracy issues resulting from human error and inconsistent data were (2020).

A web-based interface was provided by the Smart Core Price Prediction System (2021) to help farmers with financial planning by forecasting changes in crop prices. To predict price patterns over a 12-month period, the system used decision tree models.[8] Before the model was trained, data preparation methods such as feature selection and monthly rainfall analysis were used. According to the study's findings, agricultural stakeholders may make well-informed decisions by combining machine learning forecasts with an easy-to-use dashboard.

ML for Crop Yield Prediction and Analysis

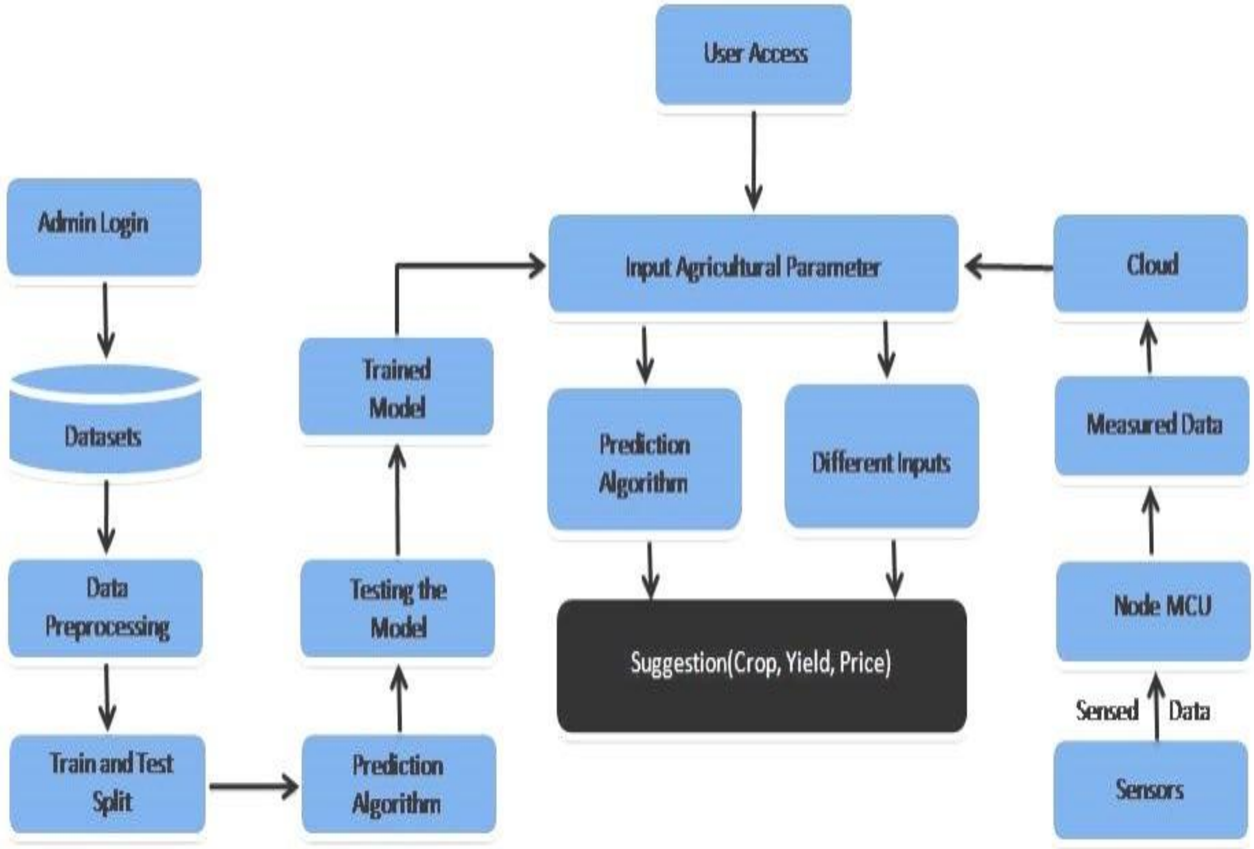
A variety of machine learning approaches have been used to enhance agricultural output forecasts. By using 10 years of historical data, Random Forest models have proven successful in predicting harvests both locally and globally. By taking into account interactions between climate, soil, and farm management, advanced regression models like Interaction Regression have improved the accuracy of soybean and maize yields. Furthermore, research shows how climatic variability, such El Niño, affects palm oil output, which enhances agricultural risk assessment. The best fertilizer application techniques for increased yields and sustainable farming methods are the main topics of recent research.[9] While studies on fertilizer placement depth show that a depth of 10 cm promotes maize development, research on maize hybrids shows that nitrogen fertilizers considerably increase grain production. For crop-specific yield estimate, ML algorithms such as ANN and Multiple Linear Regression have been used. High accuracy is demonstrated by sophisticated machine learning models like Lasso, Kernel Ridge, and Stacked Regression; some of these models can achieve error rates under 1%. Additionally, models with an R2 of 0.85 have been evaluated for yield forecasting using Decision Trees and Support Vector Regression.[10]

Proposed Method

In order to improve forecast accuracy, we present a ML based crop yield and price prediction system that makes use of the RFA and the BP Algorithm. To provide accurate forecasts, the architecture is made to methodically combine meteorological, soil, and economic factors.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 3, 2025
https://theaspd.com/index.php

Our system uses five functional modules to increase efficiency, each of which adds to the prediction process as a whole. The system technique is depicted in the image below.
Figure 2 Proposed System Architecture



Pattern Recognition and Data Analysis
Finding patterns and relationships between different parameters is the responsibility of the first module, which analyzes historical agricultural data. The RF Algorithm and Multiple Linear Regression are used in this module to forecast changes in agricultural output and pricing. Mean Absolute Percentage Error (MAPE) is castoff to assess accuracy of these forecasts and identify best forecasting strategy.

Forecasting Models

The machine learning model is trained on historical datasets in the second module, and its performance is assessed on data that hasn't been seen before. Based on user inputs, it estimates agricultural outcomes using a variety of prediction methods. To increase model accuracy and dependability, this module also uses repeated testing and parameter adjustment.[11]

Processing Results and Assisting with Decisions

After processing the gathered data, the last module provides tailored insights for agricultural decision-making. For various prediction goals, the method produces distinct outputs, including:
Crop price forecasting is the process of estimating market price trends using economic factors and historical data.
Yield estimation is the process of estimating the predicted crop yield by the analysis of agronomic, meteorological, and soil parameters.
Recommending the most appropriate crops in light of economic and environmental factors is known as optimal crop selection.
These modules are included into the suggested system to improve agricultural production, efficiency, and profitability through a data-driven decision support mechanism. To further improve precision farming methods,

future advancements will incorporate sophisticated IoT sensors, AI-powered forecasting models, and automated user assistance systems.
To improve agricultural decision-making, the suggested system architecture show in figure 2 for agricultural prediction combines predictive modeling, machine learning techniques, and data preparation. An agricultural dataset, which includes historical records of several variables including crop output, weather, soil quality, and market pricing, is the starting point for the procedure. In the data preparation step, null values are eliminated and the input data is standardized to address the missing or inconsistent values that are frequently present in these datasets.
Following preprocessing, the dataset is sent to the training and testing section, where ML models are used to find trends and produce forecasts. To increase prediction accuracy, the system makes use of a variety of ML methods, such as RF, DT, and Polynomial Regression. Every algorithm is assessed according to how well it can forecast various outcomes, including crop output, price predictions, and the best crop to choose. The accuracy of the results is then examined and contrasted.
The trained model generates forecasting findings in the last stage, known as "generating predictive insights," which are then verified using accuracy comparison measures. This guarantees that the most dependable model is chosen by the system to provide real-time agricultural suggestions. Data-driven farming decisions are made possible by the system's integration of machine learning, which enhances agricultural techniques' efficiency and production.

RESULTS AND DISCUSSION

Agricultural Dataset

Agricultural datasets including historical records on crop output, market prices, soil characteristics, and meteorological data are first gathered by the system. These datasets serve as the basis for:
Price prediction is the process of estimating agricultural market values using past patterns and affecting variables.
Crop prediction is the process of determining which crops are best to produce based on historical data, soil characteristics, and weather patterns.
Crop-Yield Prediction: Predictive modeling and historical data are used to estimate a crop's predicted yield.[13]
By improving decision-making in agriculture, this data-driven strategy maximizes profitability and production. These datasets form the foundation for predictive modeling and enable data-driven decision-making in agriculture.[14]

Table 1 Example for the Crop Recommendation Dataset

N	P	K	humidity	temperature	rainfall	ph	label
78	35	44	84.673533	26.54348	183.62257	7.0726552	rice
60	55	45	83.321291	21.405769	287.57635	5.9357454	rice
65	37	40	83.595122	23.359054	188.41366	5.3333226	rice
61	44	17	71.574769	26.100184	102.26624	6.9317565	maize
71	54	16	63.690705	22.613953	87.759857	5.7499121	maize
80	43	16	71.591368	23.558094	66.719467	6.6579653	maize
23	72	84	17.139126	19.020277	79.926081	6.9202578	chickpea
40	72	77	16.981173	17.024456	88.551143	7.4859967	chickpea
39	58	85	15.409717	17.887475	68.549919	5.9969337	chickpea
36	56	83	19.762946	18.897215	69.095477	7.4526709	chickpea
48	65	78	14.337406	17.437714	73.09204	7.8611248	chickpea
49	69	82	15.365447	18.315493	81.787463	7.2631155	chickpea
40	58	75	14.779596	18.591771	89.609451	7.1680955	chickpea
13	60	25	20.591693	17.136774	128.25862	5.6859766	kidneybeans
31	55	22	21.333114	22.913245	109.22556	5.8731794	kidneybeans
25	70	16	18.905639	19.634332	106.35183	5.7592303	kidneybeans

40	64	16	24.245875	16.433342	140.37178	5.9266785	kidneybeans

Table 2 Example for the Agricultural Crop Production and Yield Dataset

State	District	Crop Year	Crop	Season	Temperature	Soil moisture	humidity	area	Production
0	78	2000	0	1	36	35	45	1254	2000
0	78	2000	46	1	37	40	46	2	1
0	78	2000	59	1	36	41	50	102	321
0	78	2000	3	4	37	42	55	176	641
0	78	2000	12	4	36	40	54	720	165
0	78	2000	15	4	34	45	52	18168	65100k
0	78	2000	21	4	34	55	62	36	100
0	78	2000	67	4	35	50	59	1	2
0	78	2000	69	4	25	55	55	5	15
0	78	2000	70	4	36	35	45	40	169
0	78	2001	0	1	37	40	46	1254	2061
0	78	2001	46	1	36	41	50	2	1
0	78	2001	59	1	37	42	55	83	300
0	78	2001	12	4	36	40	54	719	192
0	78	2001	15	4	34	45	52	18190	64430k

Table 3 Example for the Fertilizer Recommendation Dataset

Humidity	Temperature	Moisture	Crop Type	Soil Type	Nitrogen	Phosphorous	Potassium	Fertilizer Name
52	26	38	Maize	Sandy	37	0	0	Urea
52	29	45	Sugarcane	Loamy	12	36	0	DAP
65	34	62	Cotton	Black	7	30	9	14-35-14
62	32	34	Tobacco	Red	22	20	0	28-28
54	28	46	Paddy	Clayey	35	0	0	Urea
52	26	35	Barley	Sandy	12	13	10	17-17-17
50	25	64	Cotton	Red	9	10	0	20-20
64	33	50	Wheat	Loamy	41	0	0	Urea
60	30	42	Millets	Sandy	21	18	0	28-28
58	29	33	Oil seeds	Black	9	30	7	14-35-14
54	27	28	Pulses	Clayey	13	40	0	DAP
62	31	48	Maize	Sandy	14	12	15	17-17-17
50	25	65	Cotton	Loamy	36	0	0	Urea
62	32	41	Paddy	Clayey	24	22	0	28-28

Preprocessing of Data

Before training model, preprocessing of data is performed to clean and refine the dataset. This process involves removing null values, handling missing or inconsistent data, and normalizing input features. Effective preprocessing enhances efficiency and accuracy of the predictive model. Once preprocessing is complete, the

cleaned and structured dataset is ready for further processing. The dataset is divided into training and testing subsets to ensure that the model is validated on unseen data before deployment.
The dataset consists of multiple attributes related to crop yield prediction, fertilizer recommendation, and price forecasting, including:
Crop Yield Analysis and Prediction Attributes: Name of State, Name of District, Crop Year, Crop, Season, Temperature, Soil Moisture, Humidity, Area and Production
Fertilizer Recommendation Attributes: Humidity, Temperature, Moisture, Crop Type, Soil Type, Nitrogen, Phosphorus, Potassium and Name of Fertilizer
Crop Prediction Attributes: Nitrogen(N), Phosphorus(P), Potassium(K), Humidity, Temperature, Rainfall, pH

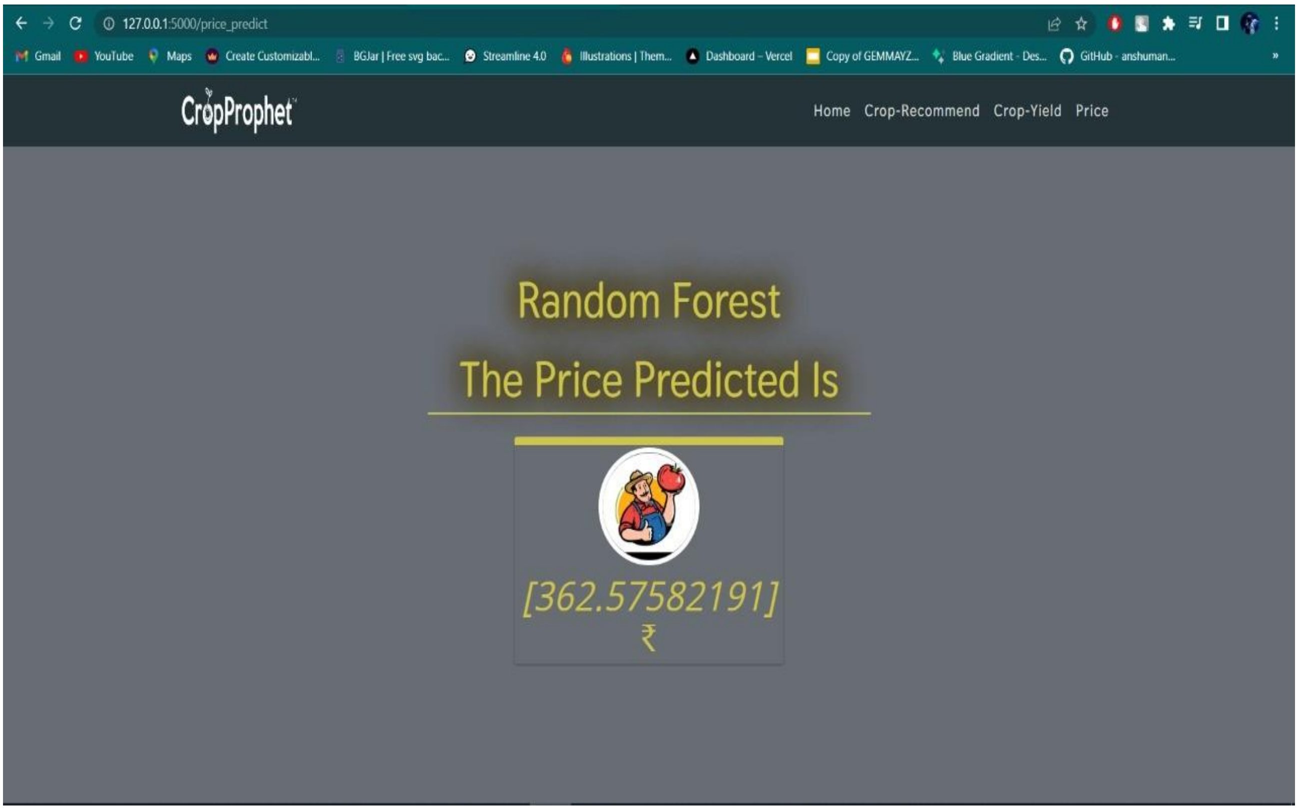
Feature Scaling & Normalization:
Min-Max Scaling: Normalize values between [0,1]:

$$Y_{scaled} = \frac{Y - Y_{min}}{Y_{max} - Y_{min}}$$

Z-Score Standardization: Normalize mean = 0, std = 1:

$$Z = \frac{X - \mu}{\sigma}$$

Model Training and Testing
In order to mitigate over fitting and uncover significant patterns from historical data, ML models are trained on a preprocessed dataset. Through 82% of data utilized for training to optimize parameters and 19% set aside for testing to verify model performance, the 80%-20% split guarantees the best possible balance between model learning and evaluation.
Stratified sampling is used to preserve class distribution across training and testing sets for categorical classification problems, reducing bias and enhancing model stability. By improving generalization, this method makes sure the model is resistant to changes in the data.
Trained model is assessed on unseen data during the testing phase to gauge its capacity for generalization and prediction accuracy. Prediction accuracy is measured using performance indicators like MAPE and RMSE, which guarantee the model's efficacy in practical applications.



Machine Learning Algorithms & Prediction Results

To improve prediction accuracy, the system employs multiple machine learning techniques. Random Forest, a powerful ensemble method, enhances accuracy by combining multiple decision trees. The Decision Tree algorithm uses a rule-based approach to partition data for informed predictions. Polynomial Regression is applied to capture non-linear relationships in crop yield and price forecasting. After training and testing, the model generates predictions for crop yield and market price. These results are used to assist farmers and stakeholders in making strategic decisions regarding crop selection, pricing, and resource allocation.

The method guarantees very accurate, comprehensible, and useful predictions by utilizing a mix of ensemble learning, rule-based partitioning, and non-linear regression; this eventually helps with risk management and strategic agricultural planning.

Figure 4 Generated Result of the Proposed Approach

Forecasting and Guidance

Following training and validation, the models produce forecasts for market price and crop production, allowing stakeholders to make informed choices. These forecasts corroborate:

Farmers: Making the best crop choices based on market trends and anticipated output.

Agricultural planners: predicting high-yield crops and possible price swings to allocate resources effectively.

Market analysts: By examining anticipated supply-demand patterns, pricing methods may be improved.

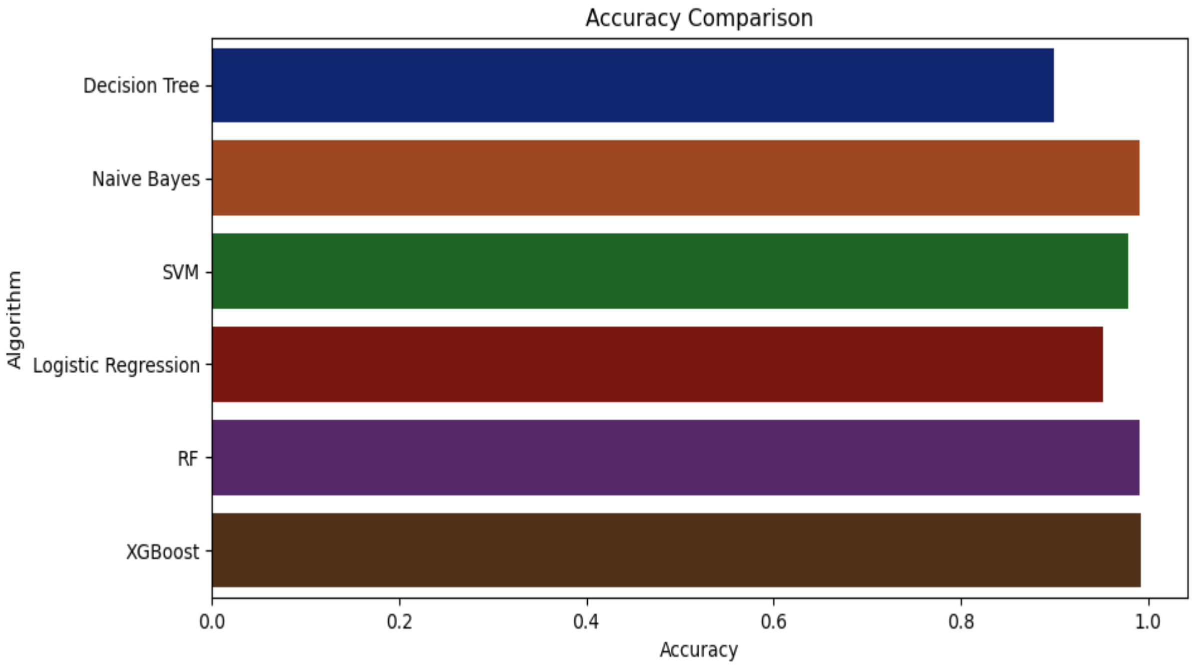


Figure 5 Model Performance Comparison proposed method and comparing other methods

Evaluation Metrics

Model performance is evaluated by comparison of predicted values with time real results. Metrics such as MAPE and RMSE are used to assess accuracy. The best-performing model is selected for implementation based on these evaluations.

Table 4. Model Performance Evaluation Table for Different Approaches

Algorithm	MAPE (%)	RMSE
Random Forest	8.12	24.5
Decision Tree	10.45	30.2
Polynomial Regression	9.25	27.8

CONCLUSION

RFA and Back Propagation Algorithm, this study offers a ML based method for forecasting crop kind, yield, and price. Our approach helps farmers make better decisions, lower financial risks, and increase agricultural output by utilizing both historical and current data. We found major obstacles to pricing and yield prediction, including climate variability and inconsistent data, by conducting a thorough analysis of the body of current work. According to experimental data, Random Forest and Back Propagation are very helpful for agricultural forecasting since they perform better than other algorithms with respect to dependability and accuracy.

The study also shows how predictive analytics may be used with rural development programs like precision farming and sericulture to shape a more sustainable and resilient agricultural environment. Farmers may get a greater understanding of market trends, soil health, and climatic patterns by integrating AI-driven agricultural forecasts with village-level development initiatives. This will eventually promote economic resilience and better crop management techniques.

REFERENCES

1. F. F. Haque, 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/WFIoT48130.2020.9221459.
2. M. Kavita and P. Mathur, "Crop Yield Estimation in India Using Machine Learning," 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 2020, pp. 220-224, doi: 10.1109/ICCCA49541.2020.9250915
3. P. S. Nishant, 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.
4. R. Kumar, 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 (ICSTM), 2015, pp. 138-145, doi: 10.1109/ICSTM.2015.7225403.
5. 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).
6. Bojtor, C.; Mousavi, S.M.N.; Illés, Á.; 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.
7. 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. (2021).
8. 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)
9. Ranjani Dhanapal et al 2021 J. Phys.: Conf. Ser. 1916 012042 Crop price prediction using supervised machine learning algorithms.
10. Jain, Ayush & Marvaniya, Smit & Godbole, Shantanu & Munigala, Vitobha. (2022). Towards Context-based Model Selection for Improved Crop Price Forecasting. 10.1145/3493700.3493725. (January 2022).
11. 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.
12. Ghutake, I., Verma, R.K., Chaudhari, R., & Amarsinh, V. (2021). An intelligent Crop Price Prediction using suitable Machine Learning Algorithm. ITM Web of Conferences.
13. 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.
14. 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.