IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 12, Issue 10, November 2025 https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2025.v1210.04

Integrated Machine Learning Models for Cotton Crop : Soil Density Analysis, Health Monitoring and Region Based Price Prediction

Prof.Vinod V Kulkarni
Computer Science and Engineering - Data Science
AMC Engineering College, Bengaluru,India
kulkarni844@gmail.com
https://orcid.org/0000-0001-6987-2657
Dhiksha CG, Kavya M, Hamsana K S, Monisha N
Computer Science and Engineering - Data Science
AMC Engineering College, Bengaluru, India
dhikshacg@gmail.com, kavyamanjunathan@gmail.com
hamsanaks0411@gmail.com, nmonisha.nanda@gmail.com

Publication History
Manuscript Reference No: IJIRAE/RS/Vol.12/Issue10/OCAE10086
Research Article| Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRAE/RS/Vol.12/Issue10/OCAE10086 Received:22,October 2025,Revised: 28,October 2025,Accepted:31,October 2025, Published Online: 21, November 2025.
https://www.ijirae.com/volumes/Vol12/iss-10/04.OCAE10086.pdf
Citation: Prof.Vinod V Kulkarni,Dhiksha,Kavya,Hamsana,Monisha (2025), Integrated Machine Learning Models for Cotton Crop : Soil Density Analysis, Health Monitoring and Region Based Price Prediction ,IJIRAE: International Journal of Innovative Research in Advanced Engineering, Volume 12, Issue 10 of 2025 pages 372-376
Doi:> https://doi.org/10.26562/ijirae.2025.v1210.04
BibTeX Key: Prof.Vinod V Kulkarni @2025 Integrated
IJIRAE papers should be cited as IJIRAE (International Journal of Innovative Research in Advanced Engineering, AM Publications, India 2025, ISSN 2349-2163, https://doi.org/10.26562/ijirae.2025.v1210.04 The journal's official abbreviation is IJIRAE. Orcid: https://orcid.org/0009-0004-9398-7488

Abstract: Cotton plays a major role in India's agricultural sector; however, farmers continue to struggle with unpredictable soil conditions, changing weather patterns, and volatile market prices. These factors directly influence the yield and profit. This study developed a machine learning framework to support farmers in making informed, data-driven decisions. The system includes two components: an Agro-Environmental Model, which evaluates soil bulk density, fertility, and vegetation health using NDVI and IMD weather data and a Market Model, which predicts regional cotton prices from APMC datasets. The outputs from both models were integrated into a unified recommendation system that indicates where cotton can be cultivated efficiently, evaluates crop health, and anticipates market value. Early testing produced encouraging results, showing that the system can assist farmers in better planning their cultivation and marketing strategies.
Keywords: Cotton crop, machine learning, agriculture, soil density, NDVI, IMD, APMC, price prediction.

I. INTRODUCTION

Agriculture is the backbone of Indian economy. Cotton is one of the most valuable cash crops. Cotton is an essential crop for the textile industry and sustains millions of farmers, particularly in states like Maharashtra, Gujarat, Telangana and Karnataka. But it is a struggle for cotton growers, who face environment and economic hazards every season. Factors such as soil fertility, erratic rainfall, and market volatility make it difficult to know in advance what the harvest will be like or how much money will come in. Farmers have typically relied on past experience and sight to decide when they should plant or sell cotton. However these methods are often not credible and do not always work when things happen quickly. Owing to technological advances, machine learning (ML) can be used to analyze volumes of agricultural data. By merging soil, weather, and market data, ML models can uncover hidden patterns that humans might overlook. This study presents a combined ML framework that connects the environmental and economic factors of cotton farming. It focuses on two key models: the first is the Agro-Environmental Model, which considers soil density, uses NDVI for health monitoring, and includes weather forecasts. The second model is the Market Model, which predicts regional prices based on APMC data. Together, these models help farmers decide when to plant, manage their crops, and sell their produce for the best profits.

II. LITERATURE SURVEY

1. Prasad et al. demonstrated the effectiveness of Random Forest regression in predicting cotton yield using long-term agro meteorological and spectral variables. Their work established the superiority of ensemble models over traditional linear methods for more precise yield forecasting.

2. Amani et al. proposed a deep learning-based soil assessment framework to optimize nutrient utilization and reduce costs in cotton cultivation, particularly when the datasets are limited. Their findings emphasized that artificial intelligence enhances both profitability and sustainable farming practices.
3. Rayasingh investigated the economic significance of cotton and highlighted the need for improved production practices that consider climatic and market constraints.
4. Maji et al. employed image-based deep learning techniques for early identification of diseases in cotton plants, demonstrating improved crop protection and yield quality.
5. Borkar used the ARIMA model to forecast production trends across major Indian states, confirming its reliability in long-term agricultural prediction.
6. Wang et al. provided a comprehensive review of deep learning applications in yield estimation, underscoring advancements in neural architectures and the integration of environmental data to achieve higher accuracy levels.
7. Jamuna et al. explored seed quality classification using machine learning models including Decision Trees and Multilayer Perceptron, proving their relevance in selecting input seeds that significantly impact final yield.
8. Mitra et al. compared several machine learning models and revealed the efficiency of Light GBM and Random Forest in predicting cotton yield by simplifying complex weather datasets into more interpretable features.

III. PROPOSED SYSTEM

The Cotton Crop Prediction System helps farmers make better decisions about growing and selling crops. This is accomplished by modeling environmental data alongside market knowledge. In traditional prediction systems, overall yield estimation and market trend analysis are considered separately, with a deficiency of decision support. Instead, the proposed approach utilizes two learning models to treat agricultural feasibility and profitability assessments as one. The Agro-Environmental Model analyses biophysical parameters that have a directly influence cotton development. It also examines the density and composition of the nutrients in the soil, among other aspects of soil fertility. To detect crop stress early on, the model also tracks the health of the vegetation using the Normalized Difference Vegetation Index (NDVI). It also collects real-time climate information, such as temperature, humidity, and unexpected rainfall, from the Indian Meteorological Department (IMD) to calculate environmental uncertainty. Predictions of cotton yield and a value for the suitability of cultivating crops in this region-time domain are provided by this model. This assessment is verified by the market model, which examines real-time APMC prices and their past trading patterns and local changes in demand-supply to predict possible selling prices at harvest times. It also provides an estimate of profit margins by estimating the relationship between production cost assumptions and yields. Both models work together within a single predictive framework. The results were combined to create actionable recommendations.

The system offers the following to farmers:

- Grow/Do-Not-Grow decision for cotton cultivation
- Expected production yield
- Anticipated selling price
- Economic advice on the best ways to engage with the market.

The ultimate decision support insights are displayed in an user-friendly dashboard designed for the farmer, which is present in various languages - English, Kannada and Hindi available to diverse set of users. This holistic method increases the income that farmers can obtain from agriculture, mitigates climate and economic risks, and promotes sustainability in farming.

A. System Architecture

The system architecture of the proposed Region-Based Cotton Crop Prediction System employs a three-layer design consisting of the user layer, application layer, and data layer. This well-defined approach maximizes the integration efficiency between predictive engines, data sources, and GUI. The User Layer was built with a ReactJS web interface, allowing farmers to input regional crop information and view model output. With an interactive dashboard, users can access the predictions of price, profitability, and crops recommendations on a real time basis. The Application Layer is implemented with a backend based on FastAPI which manages the request elaboration, data retrieval and communication to the Machine Learning engine. It handles service orchestration, data validation, pre-processing and model inference. This layer's ML engine, centralized in nature, runs both the agro environmental model and agri-market model to make predictions & cultivation suggestions.

The Data Layer integrates multiple real-time and historical datasets required for prediction modeling, including the following:

- NDVI Satellite Data for vegetation health analysis
- OpenWeather API for dynamic climatic conditions
- APMC Market Dataset for price forecasting based on region and commodity trends

The prediction outputs from the ML engine were sent back to the frontend as structured insights comprising:

- Crop recommendation decision (Grow/Do-Not-Grow)
- Price prediction for cotton crop in specific regions
- Profitability assessment for economic sustainability

This architecture ensures scalability, data reliability, and low-latency decision generation, thereby enabling farmers to receive accurate and timely agricultural advisories.

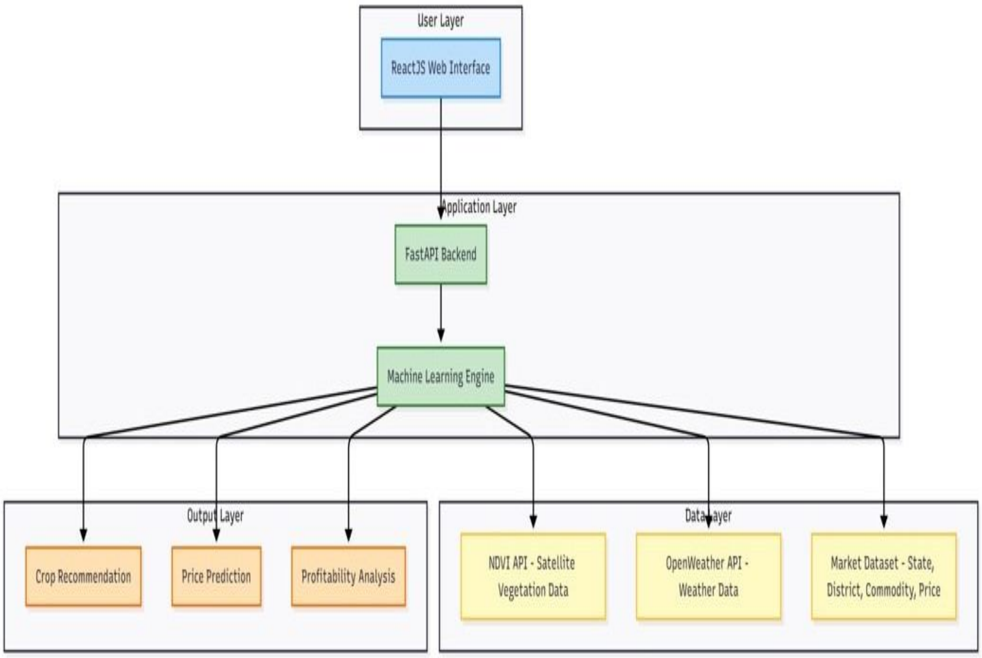

Fig. 3.1 System architecture of the proposed cotton crop prediction system integrating a ReactJS-based user interface, FastAPI backend, machine learning engine, and multi-source agricultural datasets to generate crop recommendations, price prediction, and profitability analysis.

B. Methodology

The proposed system follows a structured process to combine agricultural and market data to predict cotton crop suitability and regional price trends, which typically involves several steps. The step-by-step process is as follows:

- **Data Collection**

The research started with the acquisition of data from a number of reliable sources in order to build a diverse and balanced dataset. Soil characteristics including pH, nutrients (N, P, K), and moisture were obtained from government databases. NDVI data were utilized for the evaluation of plant health, while the combination of weather data from Open Weather and IMD APIs in real-time guaranteed up-to-date environmental information for the analysis.

- **Data Preprocessing**

The data that were collected were cleaned and aligned to maintain consistency in order to be consistent. Missing or wrong values were corrected or eliminated, and all datasets were combined into one structured table. The cleaned dataset, which acted as a filter for the model's training, was therefore highly dependable for training purposes.

- **Feature Engineering**

The whole process resulted in the design of two machine learning models one for crop suitability and the other for price prediction. Random Forest and XG Boost algorithms were selected due to their precision with intricate data. Both models went through rigorous training and assessment under metrics of accuracy and R^2 score to ultimately result in reliable predictions.

- **Model Training**

Once trained, the models were deployed to generate live predictions using current input data. The system while taking into consideration live NDVI, soil, and weather inputs not only suggests but also predicts market prices if growing cotton is successful. This integrating output results in further supporting the enhancement of both the conducting and marketing decision-making processes.

- **Prediction Phase**

The models after the training period were then put to use in the making of real-time predictions. The system while taking into consideration live NDVI, soil, and weather inputs not only suggests but also predicts market prices if growing cotton is successful. This integrating output results in further supporting the enhancement of both the conducting and marketing decision-making processes.

- **Deployment / Visualization**

Finally, the predictions and recommendations are visualized and the deployment layers are illustrated. The backend comprises the FastAPI that the system utilizes and that in turn accepts the user input, processes it, and runs the already trained models to give the output in real-time. React.js governs the front end creating a user-friendly web-based dashboard where the users can view the live results in a very clear and simple manner. The dashboard displays crop recommendations, predicted prices, and profit estimates using charts and color-coded indicators. This makes the results easy to access even for non-technical users. This stage connects technical machine learning outcomes with real-world agricultural decision making.

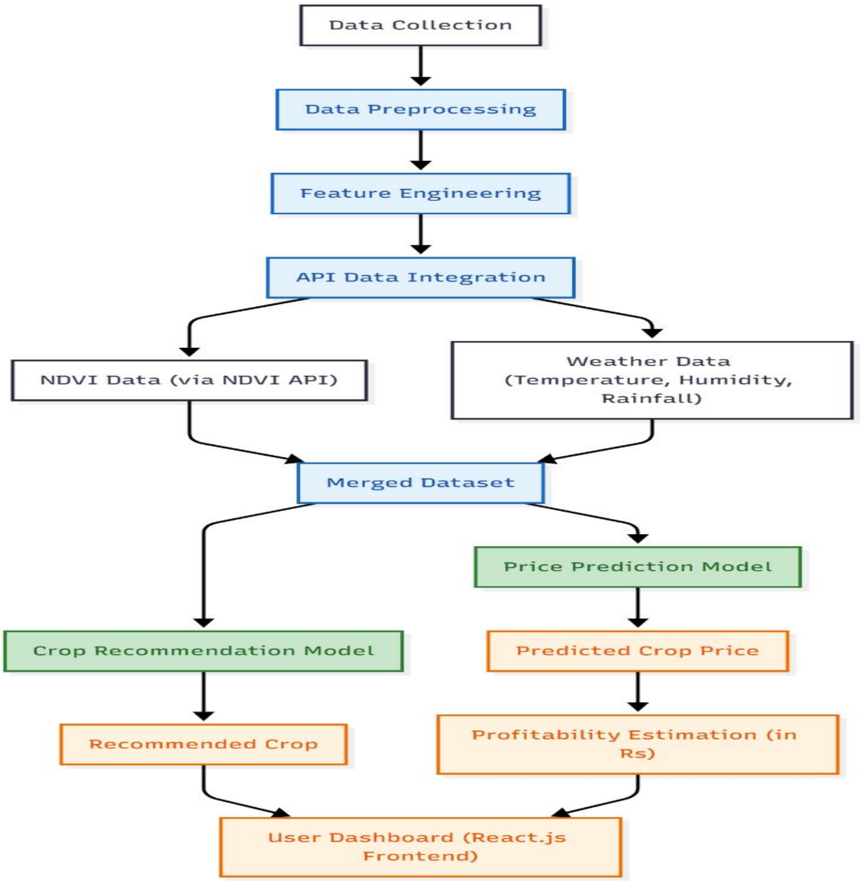

Fig 3.2: Proposed methodology flow for the cotton crop prediction system showing the process from data collection, preprocessing, and feature engineering to API integration, model training, and visualization through a React.js user dashboard.

IV. IMPLEMENTATION

The system uses a React.js-based frontend and FastAPI backend. The frontend offers a multilingual dashboard where users can view predictions and recommendations. The backend combines two machine-learning models. One model analyzed the soil, NDVI, and weather conditions. The other model predicts market prices using the APMC data. The outputs were merged to create a final recommendation, which was shown visually to help farmers make better cultivation decisions.

A. Frontend

The frontend of the Cotton Crop Prediction System was built using React.js, which supports a modular and component-driven design. React's virtual DOM and state management provide high responsiveness and optimal rendering. This is crucial for real-time agricultural decision-making systems. To create a modern and user-friendly interface, the frontend uses Tailwind CSS and Chakra UI. The Tailwind CSS provides utility-first styling for easy customization. Chakra UI offers accessible, theme-enabled UI components that improve the usability. The integration supports both light and dark modes, making it comfortable to view under different lighting conditions and on various devices. The system includes multilingual support for English, Hindi, and Kannada languages. This is done through a custom translation module that changes the UI text based on user preferences, ensuring access for farmers in different parts of India. The main interface is a dashboard that combines crop-related insights with AI-driven analytics.

The dashboard shows:

- Cotton Yield Prediction based on agricultural input parameters
- Expected Market Price to help farmers make smart economic choices
- Soil and Weather Feedback from climate and environmental datasets
- Crop Recommendations to advise whether it is best to grow or sell at a specific time.

To create interactive visualizations, the system uses Recharts, a React-based data-visualization library. This allows for the dynamic plotting of trends such as predicted yield changes and price fluctuations, making model outputs easier to understand. The frontend connects to the backend through RESTful API calls to retrieve the prediction results and external data. This asynchronous communication ensures that updates occur without the need to reload the page, boosting real-time data responsiveness.

B. Backend

The backend architecture was developed using a Flask API, which acted as a communication bridge between the user interface and machine learning models. The backend retrieves environmental and market parameters from external data sources:

- NDVI API → Vegetation index for crop growth feasibility
- OpenWeather API → Weather attributes such as rainfall, temperature, and humidity
- Historical Dataset → Soil and price-based training records

These inputs were pre-processed and engineered into meaningful feature vectors, including location (State, District), NDVI index, temperature and humidity. The backend integrates a trained machine learning model that evaluates these parameters and produces crop suitability, expected yield, market viability and finally recommends whether it grows or does not grow in the given area. Once predictions are generated, the backend processes the output into structured a JSON format and transmits it to the frontend for visualization. The system efficiently handles, prediction requests, feature preparation, result computation and secure data delivery. Thus, the backend ensures low-latency prediction, accurate computation, and continuous availability of real-time agricultural data.

V. RESULT AND DISCUSSION

The performance of the proposed Cotton Crop Prediction System was evaluated using real-time weather data, NDVI vegetation indices, and APMC market price datasets for various cotton-growing regions in India. The results demonstrate that the integrated dual-model framework, consisting of the agro-environmental model and market model, provides accurate and reliable insights for farmer decision-making. The agro-environmental model generated predictions for soil suitability and expected cotton yield based on location-specific environmental parameters. The model achieved consistent performance during testing, demonstrating strong generalization under varying weather conditions. The output included a suitability score and a corresponding Grow or Do Not Grow recommendation. The Market Model hypothesized a price range of harvested cotton on the basis of regional APMC arrivals data. The model was able to simulate changes in price trends based on seasonal demand, rainfall patterns, and market supply. The model was used in combination with the forecasted yield to compute a profit margin score for judging the economic feasibility of harvest. The system results were confirmed by comparing the predicted inputs with regional agricultural data that are publicly available. Our experimental metrics validated prediction accuracy with good error margins for practical implementation. Through visual output panels in the user interface, farmers can easily interpret the predicted yield, expected selling price, profitability estimation for optimal selling timing, and recommendation on whether to cultivate cotton. The multilingual support of the dashboard ensured that the performance results could be understood by a wider audience, including farmers with limited technical literacy.

VI. FUTURE SCOPE

The system can be extended to incorporate pest and disease surveillance using satellite-based anomaly detection and field-reported infestation data to provide timely preventive alerts. Additionally, crop rotation recommendations can be enhanced to restore soil fertility after cotton cultivation. Crops such as pigeon pea, green gram, sorghum, and groundnut can be suggested because of their nitrogen-fixing properties and ability to improve soil structure. These enhancements, along with improved climate-adaptive models will strengthen sustainability and support farmers in achieving long-term agricultural productivity.

VII. CONCLUSION

The proposed Cotton Crop Prediction System effectively integrates agro-environmental analytics and market intelligence to support farmers in making informed cultivation decisions. By utilizing soil health parameters, NDVI-based vegetation insights, and real-time weather information, the Agro-Environmental Model provides accurate yield and suitability predictions. In parallel, the Market Model analyzes APMC data to estimate price trends and profitability for region-specific cotton production. The unified predictive framework delivers actionable recommendations through an intuitive multilingual dashboard, enabling farmers to assess both environmental feasibility and economic viability prior to initiating cultivation. The system minimizes financial risk, improves resource utilization, and promotes sustainable agricultural practices. Overall, the results indicate that the proposed approach can serve as a reliable digital advisory tool to enhance decision-making in cotton farming and improve the livelihoods of agricultural communities.

REFERENCES

1. N. R. Prasad, N. R. Patel and A. Danodia, “Crop yield prediction in cotton for regional level using random forest approach,” Int. J. Agric. Sci. Res., 2023.
2. M. Rayasingh, “Exploring Cotton Crop Profile in India,” National Law University Odisha, 2022.
3. M. A. Amani and F. Marinello, “A Deep Learning-Based Model to Reduce Costs and Increase Productivity in Cotton Cultivation,” J. Agricultural Informatics, vol. 12, no. 3, 2023.
4. S. Maji et al., “Cotton Crop Certainty Identification Using Deep Learning Techniques,” in Proc. ICCCNT, Delhi, India, 2023, doi: 10.1109/CCCS.2019.8888092.
5. P. Borkar, “Time Series Analysis of Major Cotton Production States in India using Box-Jenkins Approach,” CCI Research Journal, vol. 58, pp. 1964–2022, 2023.
6. R. J. Martin et al., “XAI-Powered Smart Agriculture Framework for Enhancing Food Productivity and Sustainability,” IEEE Access, vol. 12, pp. 168412–168426, Nov. 2024, doi: 10.1109/ACCESS.2024.3492973.
7. A. Mitra et al., “Cotton Yield Prediction: A Machine Learning Approach With Field and Synthetic Data,” IEEE Access, vol. 12, pp. 101273–101288, 2024
8. Y. Wang et al., “Progress in Research on Deep Learning-Based Crop Yield Prediction,” Frontiers in Artificial Intelligence, vol. 5, pp. 1–18, 2024.