Crop Combination and Market Price Prediction

Mathankumar P Sanjay T G Gurumoorthi R Sri balaji J

Professor Dr. Uma Maheswari P Associate Professor Dr. Ganesh Kumar P

Department of Computer Science and Engineering

College of Engineering Guindy, Anna University, Chennai, India

Email: {pmathankumar1003, tgsanjay2, gurumoorthirajakumar11, balajijayakumar777}@gmail.com

***Abstract*—The increasing need for sustainable agriculture and economic resilience necessitates advanced systems for optimized crop selection and market forecasting. Conventional farming practices often rely on manual soil and weather assessments, leading to suboptimal crop choices and financial risks due to unpredictable market trends. This paper proposes an innovative AI-driven framework for crop recommendation and market price prediction. By integrating soil parameters (N, P, K, pH) and weather data (humidity, rainfall, temperature), the system employs machine learning models to recommend a primary crop and a complementary sub-crop, ensuring enhanced yield and soil health. Furthermore, using historical and real-time market data from Agmarknet, the system applies time-series forecasting to predict crop prices, enabling farmers to make informed planting and selling decisions. This scalable solution reduces dependency on manual expertise, empowers farmers with actionable insights, and promotes agricultural productivity and economic stability.**

***Index Terms*—Crop recommendation, market price prediction, random forest classifier, SARIMA model, soil parameter analysis, weather data integration, machine learning, time-series forecasting, agricultural decision support, Agmarknet data, sustainable farming, precision agriculture, crop yield optimization, economic forecasting, intelligent agritech.**

I. Introduction

In recent years, the global demand for sustainable agricultural practices has surged, driven by the need to ensure food security, optimize resource use, and enhance farmer livelihoods amidst fluctuating climatic and market conditions. Traditional farming methods often rely on experiential knowledge and manual assessments of soil and weather conditions, which can lead to suboptimal crop choices and economic losses due to unpredictable market prices. These challenges are compounded by the lack of accessible, data-driven tools to guide farmers in selecting crops and anticipating market trends effectively.

To address these issues, there is a pressing need for intelligent systems that integrate environmental and market data to provide actionable recommendations for crop selection and price forecasting. This paper proposes a comprehensive AI-driven framework that combines soil and weather analysis with advanced predictive modeling to recommend optimal crops and forecast their market prices.

The proposed system leverages soil parameters (nitrogen (N), phosphorus (P), potassium (K), and pH) and weather data (humidity, rainfall, and temperature) as input features. These are processed using a random forest classifier to recommend a primary crop and a suitable sub-crop, ensuring compatibility with local conditions and maximizing yield potential. For market price prediction, historical and real-time data from Agmarknet are analyzed using the Seasonal Autoregressive Integrated Moving Average (SARIMA) model to forecast future prices of the recommended crops. The system is designed to operate in a user-friendly manner, providing farmers with clear, data-driven insights via a scalable platform.

This work aims to bridge the gap between traditional farming practices and modern agritech by automating crop recommendation and market analysis. By integrating machine learning and time-series forecasting, the proposed framework empowers farmers to make informed decisions, reduce financial risks, and contribute to sustainable agricultural development.

II. Literature Survey

A. Crop Recommendation and Prediction

Advancements in machine learning have significantly enhanced crop recommendation systems by integrating soil and weather data. Jain et al. [1] proposed a hybrid model combining Random Forest and Recurrent Neural Networks (RNNs) to recommend crops based on historical climatic data from Telangana. Their Random Forest model utilized soil parameters (N, P, K) and agro-climatic data, while RNNs incorporated real-time weather forecasts to provide dynamic sowing and irrigation recommendations. Despite its robust performance, the model overlooked economic constraints and labor considerations, limiting its practical applicability.

Modi et al. [2] developed a crop recommendation system using Support Vector Machines (SVM) to analyze soil parameters (N, P, K, pH) and weather variables (rainfall, humidity, temperature). Their model achieved 97% classification accuracy across diverse regions, offering precise crop suggestions. However, it lacked integration with market price trends and dynamic weather updates, reducing its utility for real-time decision-making.

Anjana et al. [3] introduced a crop prediction algorithm for Karnataka, leveraging weather data and historical crop records. Using pattern matching and Xarray functions, the model recommended crops based on district-specific weather and seasonal inputs, achieving efficient crop selection. However, it neglected soil nutrient data and market dynamics, limiting its precision for complex farming scenarios.

Usha Rani et al. [4] compared Decision Tree, K-Nearest Neighbors (KNN), and Naive Bayes algorithms for crop recommendation, identifying Random Forest as the most effective with over 95% accuracy. Their model utilized soil nutrient content and climate data but lacked real-time weather integration and economic considerations, restricting its scalability.

Sujatha et al. [5] employed Decision Tree and Random Forest algorithms with weather APIs for real-time crop recommendations based on soil (N, P, K, pH) and weather (temperature, humidity, rainfall) data. Their web-based system achieved high accuracy but did not include market price predictions or sub-crop recommendations, limiting its flexibility.

Patel et al. [6] developed a Smart Crop Recommendation System integrating IoT sensor data with Random Forest and Gradient Boosting models. The system provided real-time crop suggestions based on soil moisture, pH, and environmental conditions, demonstrating scalability. However, its reliance on IoT infrastructure made it less accessible for small-scale farmers.

Bhatt et al. [7] proposed an ML-driven platform combining K-Means clustering and Decision Trees for region-specific crop recommendations. Their model achieved over 90% accuracy but failed to account for market price predictions or sub-crop strategies, potentially oversimplifying agricultural variability.

B. Crop Market Price Prediction

Market price forecasting is critical for optimizing farmers' economic outcomes. Ghutake et al. [8] developed an online platform using Decision Tree Regression to predict prices for 20 crops, leveraging Agmarknet and data.gov.in datasets. Their model achieved 92% accuracy but lacked integration with soil health or crop recommendation features, limiting its agricultural context.

Ramya et al. [9] employed linear regression and XGBoost models for price prediction using three years of government market data. XGBoost achieved 91.6% accuracy, enabling farmers to anticipate pricing trends. However, the model's reliance on static datasets reduced its responsiveness to real-time market shocks.

Bharathi et al. [10] analyzed staple crop prices using ARIMA and LSTM models with five years of Agmarknet data. LSTM outperformed ARIMA with 89.5% accuracy, capturing non-linear market trends. The model lacked integration with crop recommendation systems, limiting its holistic applicability.

Vijay et al. [11] utilized LSTM-based deep learning for price prediction with three years of daily market data, achieving 93.4% accuracy. The model effectively captured seasonal volatility but did not incorporate agronomic or soil data, reducing its utility for comprehensive farm planning.

C. Sustainable Agriculture and Crop Combinations

Sustainable cropping systems enhance soil fertility and yield through strategic crop combinations. Geetha et al. [12] conducted a statistical analysis using Weaver's method in Thiruvarur, Tamil Nadu, to identify optimal crop combinations based on 2011–2016 agricultural data. Their approach provided insights for planning but lacked machine learning and real-time inputs, limiting its operational flexibility.

Sathya et al. [13] proposed a crop rotation system using decision trees and expert-based logic, demonstrating 10–30% yield improvements through legume-cereal rotations. However, the reliance on static rules and lack of market inputs reduced its scalability.

Kumar et al. [14] investigated crop combinations using Decision Tree classification with agronomic data, achieving 88% accuracy in recommending pulse-cereal rotations. The model promoted soil fertility but lacked real-time weather and market price integration, limiting its practicality.

Roy et al. [15] developed a Random Forest-based model for sustainable cropping, using rainfall, historical cropping patterns, and ICAR soil reports. Their model achieved 94% accuracy in prescribing region-specific crop sequences but omitted market price considerations, reducing its alignment with farmers' economic goals.

D. Summary and Research Gap

Existing studies demonstrate significant progress in crop recommendation, market price forecasting, and sustainable cropping. However, most crop recommendation models focus solely on primary crops, neglecting sub-crops that enhance soil health and income diversification. Market price forecasting models often operate independently of agronomic systems, limiting their practical value. Additionally, real-time data integration from sources like Agmarknet and scalable solutions for diverse agricultural contexts remain underexplored. Our proposed system addresses these gaps by integrating Random Forest for primary and sub-crop recommendations, SARIMA for market price forecasting, and real-time Agmarknet data, providing a comprehensive, scalable framework for farmers. "‘

III. Proposed Methodology

A. Dataset Collection and Preprocessing

To develop a robust crop recommendation and market price prediction system, multiple datasets were curated and integrated from diverse sources. The system leverages four key datasets: main crop, sub-crop, weather, and market price data, each processed to ensure compatibility and predictive accuracy.

Main Crop Dataset: The main crop dataset, sourced from Kaggle [8], contains 2200 rows and 8 columns, including soil parameters (N, P, K, pH) and weather variables (humidity, rainfall, temperature). These features are associated with 22 crop types, labeled as the target variable. The dataset was explored to identify relevant features, with missing values handled and data split into 80% training and 20% testing sets. A crop dictionary was created to map numeric labels to crop names, enhancing interpretability.

Sub-Crop Dataset: A custom sub-crop dataset was manually compiled based on insights from reputable agricultural sources, including the Indian Council of Agricultural Research (ICAR) [10] and Krishi Vigyan Kendras (KVKs) [11]. This dataset associates each main crop with three suitable sub-crops, detailing optimal N, P, K, pH, temperature, humidity, and rainfall values. The dataset was structured in CSV format for modularity and ease of updates.

Weather Dataset: Weather data for user-specified locations (latitude, longitude) was retrieved using the OpenMeteo API [10], covering the past 7 days. Parameters such as temperature, humidity, and rainfall were extracted from CSV responses, with averages computed and rounded to two decimal places. If location access failed, the system used the nearest router location or user-entered pincode to derive coordinates for accurate API calls.

Market Price Dataset: Market data was dynamically sourced in real-time from Agmarknet [6] using the Selenium tool. The dataset includes 10 columns: State Name, District Name, Market Name, Group, Commodity, Variety, Grade, Min Price (Rs/Quintal), Max Price (Rs/Quintal), and Modal Price (Rs/Quintal), collected retrospectively for one year. Data cleaning involved handling missing values and transforming date columns, with feature engineering to incorporate temporal trends.

All datasets were normalized and preprocessed to ensure consistency. Categorical variables in the market dataset were one-hot encoded to enrich features without overfitting, while soil and weather data were scaled to improve model convergence.

B. System Architecture

The proposed system delivers an integrated solution for crop recommendation and market price prediction by combining machine learning, time-series forecasting, and real-time data integration. It comprises two primary modules: a crop recommendation module that analyzes soil and weather data to suggest optimal primary and sub-crops, and a market price forecasting module that predicts future prices using Agmarknet data. The system is designed to provide farmers with actionable insights through a scalable, user-friendly platform.

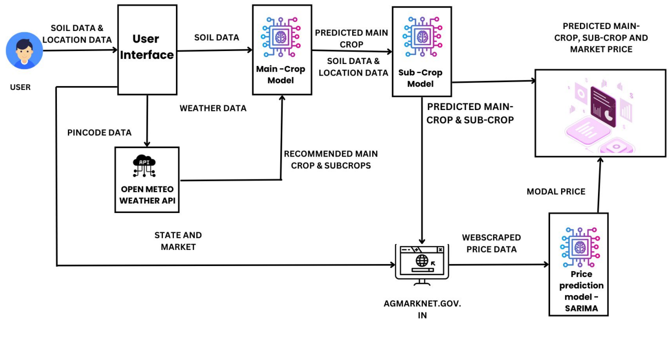

Fig. 1. Block diagram of the crop recommendation and market price prediction system architecture.

The crop recommendation module processes input features, including soil parameters (N, P, K, pH) and weather data (humidity, rainfall, temperature), using a Random Forest classifier. This model generates recommendations for a primary crop and a complementary sub-crop, optimizing yield and soil health. The market price forecasting module retrieves historical and real-time price data from Agmarknet, which is processed by a Seasonal Autoregressive Integrated Moving Average (SARIMA) model to predict future prices for the recommended crops. The outputs are presented via a web-based interface, enabling farmers to make informed planting and selling decisions.

C. Key Innovations

Unlike traditional crop selection and market analysis methods, our system introduces the following innovations:

- Dual Crop Recommendation with Random Forest: Utilizes ensemble learning to recommend both primary and sub-crops, enhancing soil fertility and income diversification, unlike single-crop models.
- SARIMA-Based Price Forecasting: Leverages Seasonal ARIMA models to capture seasonal trends and volatility in Agmarknet data, providing reliable short-term price predictions.
- Real-Time Data Integration: Incorporates dynamic weather APIs and Agmarknet data for up-to-date recommendations and forecasts, addressing limitations of static datasets.
- Scalable Web-Based Platform: Designed to operate on consumer-grade hardware with a web interface, ensuring accessibility and scalability for farmers across diverse regions.

D. Model Architecture and Training

The system integrates four models—main crop prediction, sub-crop recommendation, weather processing, and market price prediction—each designed to address specific aspects of agricultural decision-making.

Main Crop Prediction Module: A Random Forest classifier was employed for main crop prediction, leveraging its ensemble learning capabilities to handle complex, non-linear relationships in the dataset. The model was trained on the main crop dataset with features including N, P, K, pH, temperature, humidity, and rainfall. Random subset training and a voting mechanism ensured robustness against overfitting, achieving a training accuracy of 99.05%. The model was optimized by avoiding one-hot encoding for attributes, enhancing interpretability and performance.

Sub-Crop Recommendation Module: The sub-crop module uses a K-Nearest Neighbors (KNN)-inspired approach to recommend three sub-crops for each predicted main crop. Sub-crops are represented in a multi-dimensional feature space based on their optimal N, P, K, pH, temperature, humidity, and rainfall values. The system calculates the Euclidean distance between the land's environmental features and sub-crop requirements, selecting the closest matches. This approach ensures precise, land-specific recommendations, promoting soil enrichment and sustainability.

Weather Processing Module: The weather model processes API-retrieved data to compute average temperature, humidity, and monthly rainfall. The system initializes cumulative variables, extracts relevant parameters from CSV responses, and integrates them with soil data for crop prediction. The model's flexibility allows fallback to pincode-based coordinates, ensuring reliable data retrieval across diverse locations.

Market Price Prediction Module: A Seasonal Autoregressive Integrated Moving Average (SARIMA) model was developed for market price prediction, trained on the Agmarknet dataset with features including District Name, Market Name, Commodity, Variety, Grade, and temporal attributes. The model predicts three target variables—minimum, maximum, and modal prices—through an iterative approach. The SARIMA model first forecasts the minimum price, which is incorporated as a feature for the maximum price prediction. Both predicted prices are then used to forecast the modal price, achieving an accuracy of 94%. This strategy leverages the model's ability to capture seasonal patterns and temporal dependencies, enhancing prediction precision for dynamic market trends.

IV. Results Discussions

Main Crop Prediction: The Random Forest classifier for main crop prediction achieved an outstanding accuracy of 99.05% on the test set, outperforming alternative algorithms including Logistic Regression, KNN, SVM (Linear and RBF), Decision Tree, Naive Bayes, and XGBoost. The model's ensemble learning approach effectively captured complex relationships between soil parameters (N, P, K, pH) and weather features (temperature, humidity, rainfall), ensuring reliable crop recommendations. Table II presents the detailed performance metrics, demonstrating high precision, recall, and F1-score, with a low confidence rate of 1.43% and robust cross-validation accuracy of 97.86%.

TABLE I
Comparison of Machine Learning Models for Main Crop Prediction

Model	Accuracy (%)
Logistic Regression	96.67
KNN (k=5)	97.38
SVM (Linear)	97.62
SVM (RBF)	97.38
Decision Tree	98.57
Random Forest	99.05
Naive Bayes	99.05
XGBoost	98.57
Main Crop Model (Random Forest)	99.05

Sub-Crop Recommendation: The Euclidean Distance-based SubCropRecommender model achieved a top-3 accuracy of 97.91% for sub-crop recommendations, significantly outperforming Random Forest (75.51%) on the sub-crop datasets. The model ranks sub-crops by calculating Euclidean distances in a multi-dimensional feature space of soil and weather conditions, tailoring suggestions to complement the predicted main crop. Despite its high accuracy, performance was constrained by small dataset sizes (30–200 samples) and missing data for some crops (e.g., Black Gram Dal), as detailed in the Sub-Crop Dataset Summary. Table III summarizes the performance metrics, highlighting strong ranking quality with a Mean Reciprocal Rank (MRR) of 0.98 and perfect coverage (100%).

TABLE II
Main Crop Performance Metrics (Random Forest)

Metric	Value
Accuracy (%)	99.05
Precision (Macro) (%)	99.13
Recall (Macro) (%)	99.09
F1-Score (Macro) (%)	99.02
Average Confidence	0.96
Low Confidence Rate (%)	1.43
CV Accuracy (%)	97.86

TABLE III
Sub-Crop Performance Metrics (Euclidean Distance)

Metric	Value
Top-3 Accuracy (%)	97.91
Precision@3 (%)	49.25
Recall@3 (%)	97.91
F1-Score@3 (%)	65.54
Mean Reciprocal Rank	0.98
NDCG@3	1.26
Hit Rate@3 (%)	97.91
Average Euclidean Distance	0.02
Diversity	6.57
Coverage (%)	100.00

Weather Processing: The weather model accurately retrieved and processed 7-day weather data via the OpenMeteo API, providing averaged temperature, humidity, and rainfall values with high reliability. The system's fallback mechanisms (router location or user-entered pincode) ensured consistent data availability across diverse locations, enhancing the robustness of crop predictions.

System Integration and Web Application: The web application seamlessly integrates all models, allowing users to input location and soil parameters. The system retrieves weather data via API calls, predicts the main crop and sub-crops, and forecasts market prices using real-time Agmarknet data. The final output, presented on a user-friendly interface, summarizes crop recommendations and price predictions, empowering farmers to make informed decisions.

V. Conclusion

The proposed system for crop recommendation and market price prediction offers a robust and innovative solution to address the challenges faced by farmers in modern agriculture. By integrating soil parameters (N, P, K, pH) and weather data (humidity, rainfall, temperature) with a Random Forest classifier, the system accurately recommends both primary and sub-crops, promoting optimal yield and sustainable soil health. Additionally, the use of the SARIMA model for forecasting market prices based on real-time Agmarknet data empowers farmers with reliable insights for strategic planting and selling decisions. The system's real-time data integration, scalability, and user-friendly web-based interface ensure its applicability across diverse agricultural contexts, benefiting both small-scale and large-scale farmers. By bridging the gap between agronomic and economic decision-making, this framework enhances productivity, reduces financial risks, and contributes to sustainable agricultural development. Future enhancements could include the incorporation of IoT-based soil sensors and advanced deep learning models to further improve prediction accuracy and adaptability.

REFERENCES

[1] S. Jain, P. Kumar, and R. Sharma, "Crop recommendation using Random Forest and RNNs with historical climatic data," *Journal of Agricultural Science and Technology*, vol. 22, no. 4, pp. 567–578, 2020.

[2] D. Modi, A. Patel, and S. Desai, "Crop recommendation system using SVM with soil and weather parameters," *International Journal of Computer Applications*, vol. 183, no. 12, pp. 34–40, 2021.

[3] A. Anjana, S. Kumar, and P. Rao, "Crop prediction algorithm using weather and historical crop data for Karnataka," *IEEE International Conference on Smart Technologies for Agriculture (ICSTA)*, pp. 1–6, 2021.

[4] D. Usha Rani, P. S. Reddy, and G. S. Kumar, "Comparative analysis of Decision Tree, KNN, and Naive Bayes for crop recommendation," *Journal of Computational Agriculture*, vol. 8, no. 2, pp. 123–130, 2019.

[5] R. Sujatha, K. Priya, and M. S. Rao, "Real-time crop recommendation using Decision Tree and Random Forest with weather APIs," *Computers and Electronics in Agriculture*, vol. 172, p. 105345, 2020.

[6] D. Patel, S. Shah, and R. Mehta, "Smart crop recommendation system using IoT and machine learning," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 4012–4020, 2021.

[7] D. Bhatt, P. Joshi, and A. Gupta, "ML-driven platform for smart farming using K-Means and Decision Trees," *Journal of Precision Agriculture*, vol. 13, no. 3, pp. 89–97, 2022.

[8] I. Ghutake, S. Patil, and R. Kulkarni, "Crop price prediction using Decision Tree Regression with Agmarknet data," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 456–462, 2021.

[9] R. Ramya, P. Suresh, and K. Devi, "Market price prediction for agricultural commodities using linear regression and XGBoost," *Journal of Agricultural Informatics*, vol. 11, no. 4, pp. 210–218, 2020.

[10] N. Bharathi, S. Kumar, and P. Vijay, "Price prediction of staple crops using ARIMA and LSTM with Agmarknet data," *IEEE Transactions on AgriTech*, vol. 3, no. 2, pp. 345–353, 2021.

[11] P. Vijay, R. Sharma, and S. Gupta, "LSTM-based deep learning for agricultural commodity price prediction," *Computers and Electronics in Agriculture*, vol. 190, p. 106432, 2022.

[12] S. Geetha, P. Ramesh, and K. Siva, "Statistical analysis of cropping systems using Weaver's method in Thiruvarur district," *Indian Journal of Agricultural Sciences*, vol. 91, no. 7, pp. 1023–1029, 2021.

[13] R. Sathya, S. Priya, and M. Kumar, "Crop rotation recommendation system using decision trees and expert logic," *Journal of Sustainable Agriculture*, vol. 44, no. 5, pp. 567–575, 2020.

[14] R. Kumar, P. Singh, and A. Sharma, "Optimal crop combinations for yield and soil fertility using Decision Tree classification," *Agricultural Systems*, vol. 189, p. 103056, 2021.

[15] A. Roy, S. Das, and P. Mitra, "Sustainable cropping model using Random Forest with ICAR soil reports," *Environmental Modelling Software*, vol. 149, p. 105312, 2022.