

An Integrated Decision-Support Framework for Crop Selection Using Suitability Prediction, Market Price Forecasting for Profit Optimization

Khan Ayesha Muhammed Irfan* Dr. Manoj Singh**

*Department of Computer Science (SIES College of Arts, Science and Commerce Sion, Mumbai, India)
(Email: kayesha6161@gmail.com)

**Department of Computer Science (SIES College of Arts, Science and Commerce Sion, Mumbai, India)
(Email: manojks@sies.edu.in)

Abstract:

Farmers typically select crops based on either agronomic suitability or market trends, rarely both together, which can lead to growing a well-suited but comparatively low-profit crop. This paper presents a four-stage system that first predicts crop suitability from soil nutrients (N, P, K), pH, soil moisture, temperature, humidity and rainfall using a Random Forest classifier trained on a public 22-crop dataset (99.32% held-out accuracy), then forecasts market price for every agronomically suitable candidate crop using separate per-crop regression models trained on historical Government of India Agmarknet mandi price records, then computes expected monthly profit ($\text{yield} \times \text{price} - \text{cultivation cost}$, normalized by crop growth duration) rather than ranking candidates by price alone as in prior work, and finally surfaces a forecast-reliability risk label and an interactive what-if scenario tool. A case study demonstrates the system recommending cotton over the more agronomically suitable maize (72.7% vs. 27.0% suitability) due to a substantially higher expected monthly profit, illustrating the central contribution: that suitability-only and price-only recommendation strategies can diverge from the profit-maximizing choice, and that explicitly modeling cultivation cost and crop duration closes a gap left open in recent hybrid suitability-price literature.

Keywords — crop recommendation, market price forecasting, Random Forest, profit optimization, risk-aware decision support.

I. INTRODUCTION

Crop selection is one of the highest-stakes decisions a farmer makes each season, yet it is typically made using one of two disconnected sources of information: agronomic knowledge of what grows well in the local soil and climate, or recent market experience of what has sold for a good price. These two signals are rarely combined systematically. A crop that is agronomically optimal is not necessarily the most profitable choice once cultivation cost, yield, and prevailing or forecasted market price are accounted for, and a crop with an attractive forecasted price may not be viable in the farmer's actual soil and weather conditions at all.

This paper presents a system that treats crop recommendation as a profit-optimization problem constrained by agronomic suitability, rather than a suitability-only or price-only prediction problem. The system is implemented end-to-end, including a trained suitability classifier, per-crop price forecasting models trained on real government market data, a duration-normalized profit engine, a forecast-reliability risk

indicator, and an interactive what-if scenario tool, delivered through a web interface.

II. PROBLEM STATEMENT

Given a farmer's soil parameters and expected seasonal weather conditions, the problem is to recommend the crop, among all agronomically suitable candidates, that maximizes expected monthly profit. This is framed as a two-stage prediction problem (suitability classification, followed by price regression) combined with a deterministic economic decision rule (profit computation and ranking), rather than a single end-to-end model, so that each stage can be independently validated and explained.

III. OBJECTIVES

The objectives of this work are to:

- (1) train a multi-class suitability classifier across soil and weather features, including soil moisture.
- (2) train independent price forecasting models per crop using real historical market data.
- (3) design a profit formula that fairly compares crops of different growth durations.

- (4) quantify and surface forecast reliability rather than presenting point estimates as certainties.
- (5) provide an interactive mechanism for exploring how the recommendation responds to changing price or weather assumptions.
- (6) deploy the complete pipeline as a usable web application.

IV. RELATED WORK

Crop recommendation research mainly focuses on agronomic suitability, market-price forecasting, or a combination of environmental and economic factors. Senapaty et al. [1] developed a machine-learning-based crop recommendation system using multiple classification algorithms and agricultural data. Similarly, Mahale et al. [6] combined LSTM-based weather forecasting with Random Forest for crop recommendation in Maharashtra. These studies primarily focus on crop suitability and recommendation based on agricultural and environmental information.

Market-price forecasting is another important research direction. Mahajan et al. [7] investigated machine-learning methods for forecasting agricultural commodity prices using historical market data. Although such models provide useful price predictions, price forecasting alone does not determine whether a crop is suitable for a particular soil and weather condition.

Several studies have attempted to combine crop suitability with economic information. Palle and Raut [2] proposed a crop recommendation system using weather attributes, soil properties, and crop prices. Daniel et al. [3] introduced RANC-CROP, combining soil-based crop recommendation with price estimation. More recently, Sindhur et al. [4] proposed a hybrid framework combining Random Forest-based crop suitability with LSTM-based market-price forecasting. Sam and D'Abreo [5] also explored crop recommendation using environmental and economic factors. These studies demonstrate the potential of integrated crop and market analysis, but they do not explicitly use the duration-normalized net-profit formulation adopted in this work.

Explainable AI has also been applied to crop recommendation. Shams et al. [8] proposed XAI-CROP and used LIME to provide interpretable explanations for crop recommendations. Turgut et al. [9] developed AgroXAI using ELI5, LIME, and SHAP for local and global explanations, along with counterfactual explanations for alternative crop recommendations.

Based on the reviewed literature, existing systems address crop suitability, price forecasting, economic recommendation, or explainability, but their integration remains limited. The proposed work combines agronomic suitability prediction, market-price forecasting, yield and cultivation-cost-based profit calculation, duration

normalization, forecast-reliability risk assessment, and interactive what-if analysis within a single decision-support pipeline. This enables the system to recommend not only a suitable crop but also the comparatively more profitable option among suitable candidates.

V. DATASET

A. Soil and Weather Data

The soil and weather data for 22 crops was obtained from a public crop recommendation dataset, containing 2,200 instances with features including nitrogen, phosphorus, potassium, temperature, humidity, pH, rainfall and Soil moisture. The provenance of the soil-moisture feature should be documented according to the specific dataset version used.

B. Market Price Data

Historical monthly market-price data (₹/quintal) was collected from Government of India Agmarknet records for the selected crops. The data was used to develop crop-wise market-price forecasting models and support the economic comparison of suitable crops.

C. Profitability Parameters

Yield (quintal/ha), cultivation cost (Rs./ha) and crop duration (months) were taken from published government and academic cost-of-cultivation sources where available, and estimated otherwise. These values are fixed reference values per crop and are not predicted by the model. Crop duration (months from sowing to harvest) was taken from FAO and TNAU crop production references. For perennial and tree crops, a 12-month cycle is used to represent one annual harvest."

VI. METHODOLOGY

A. Crop Suitability Prediction

A Random Forest classifier (300 trees) is trained on the 8-feature input (N, P, K, pH, soil moisture, temperature, humidity, rainfall) against the 22-crop label, using an 80:20 stratified train-test split.

Rather than returning only the single top-predicted crop, the system retains every crop whose predicted probability exceeds a 5% threshold as a candidate, since more than one crop is frequently viable for the same conditions — this candidate set, not a single top-1 prediction, is what subsequent stages evaluate for profitability.

B. Market Price Forecasting

For each crop independently, a lag-feature regression problem is constructed from its own monthly price history: the preceding three months' prices serve as input features to

predict the next month's price. A separate Random Forest Regressor (100 trees, max depth 4) is trained per crop using a chronological (non-shuffled) 80:20 train-test split, to avoid future information leaking into training, which would occur under a random split. Forecasts more than one month ahead are produced recursively, feeding each prediction back in as a lag input for the next step.

C. Profit-Based Recommendation

For every candidate crop, expected monthly profit is computed as:

```
profit_per_month = (yield_per_hectare ×
forecasted_price - cultivation_cost) /
crop_duration_in_months
```

Normalizing by crop duration ensures that a fast-cycling annual crop and a slow perennial crop are compared on a like-for-like basis return per month the land is occupied rather than by raw per-cycle profit, which would systematically favor long-duration, high-yield crops regardless of how long they tie up the land. The candidate with the highest monthly profit is recommended.

D. Forecast Reliability (Risk Assessment)

Each crop's price model carries a backtested Mean Absolute Percentage Error (MAPE) from its chronological test split. This is converted into a three-level risk label (Low: MAPE < 5%; Medium: 5–15%; High: > 15%), shown alongside every recommendation, so that a user can distinguish a profit figure built on a reliable forecast from one built on a highly uncertain one.

E. What-If Scenario Exploration

A lightweight rule-based natural-language interface accepts single-factor hypothetical questions (e.g., “what if price drops by 20 percent”) and re-runs the recommendation pipeline under the adjusted assumption, displaying the original and adjusted recommended crop side by side. Compound multi-factor questions are detected and rejected with a clarifying prompt rather than silently misattributing a single parsed value to multiple factors.

VII. SYSTEM ARCHITECTURE

The pipeline follows four sequential stages:

- (1) the user supplies soil and weather readings.
- (2) the suitability classifier returns a probability-ranked set of candidate crops.
- (3) each candidate's price is forecasted by its own regression model.
- (4) expected monthly profit and a risk label are computed per candidate, and the highest-profit candidate is presented as the recommendation, alongside the full ranked candidate table and an interactive what-if tool.

The system is deployed as a multi-page Streamlit web application with Home, How It Works and About sections.

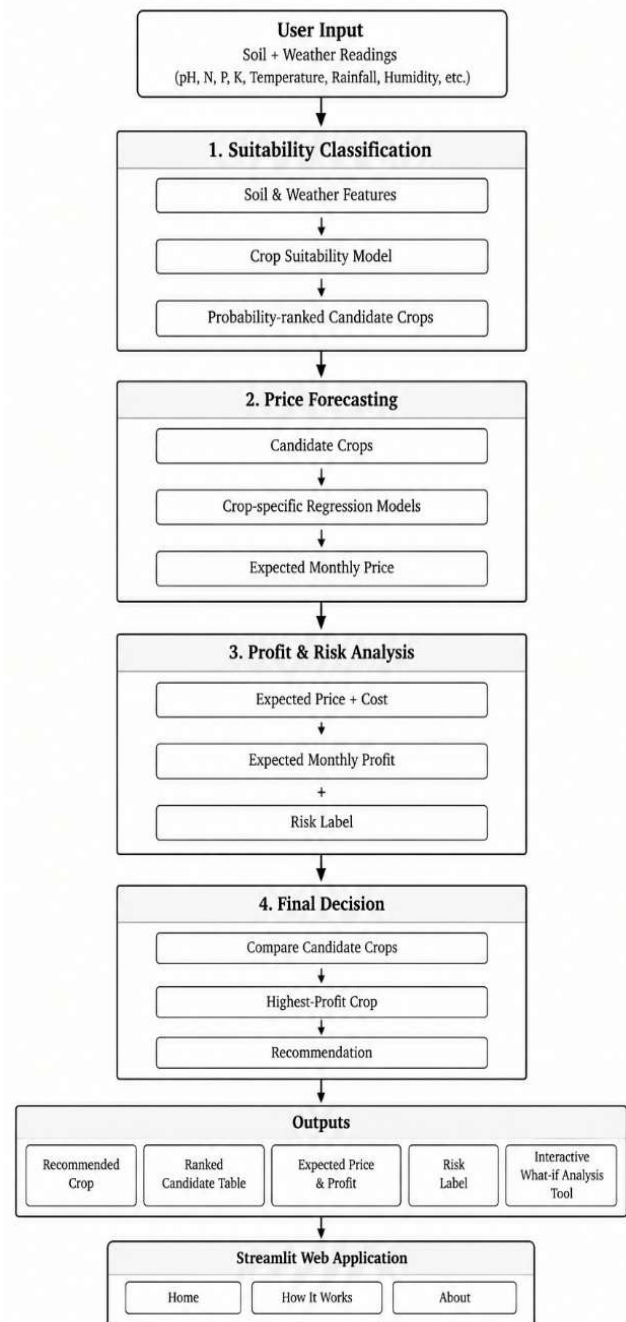


Fig. 1 SYSTEM ARCHITECTURE

VIII. EXPERIMENTAL RESULTS

A. Suitability Model Performance

On the held-out 20% test split, consisting of 440 out of 2,200 instances, the Random Forest suitability classifier

achieved an overall accuracy of 99.32% across the 22 crop classes.

B. Price Forecasting Performance

The price forecasting performance of the crop-specific regression models was evaluated using Mean Absolute Percentage Error (MAPE) on a chronological test split. The MAPE values for all 22 crops are presented in Fig. 2

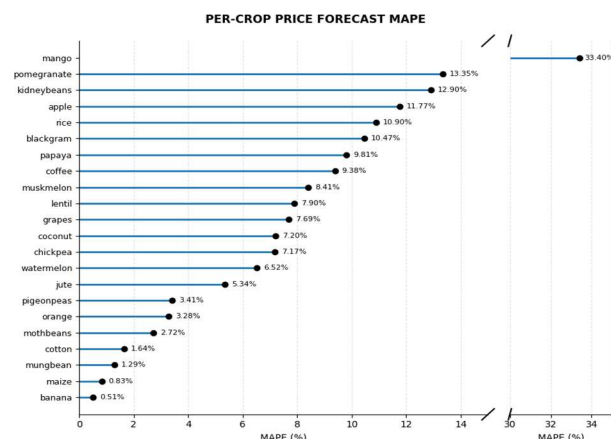


Fig. 2 PER-CROP PRICE FORECAST MAPE

Sixteen of the 22 crops achieved MAPE below 10%. The elevated MAPE for mango (33.40%) and pomegranate (13.35%) is attributed primarily to the small number of chronological test observations available (approximately 2 months per crop, given the 13-month historical window), which inflates per-crop MAPE variance rather than necessarily indicating poor model fit; this is discussed further in Section XII.

IX. FEATURE IMPORTANCE

Table I reports Random Forest feature importances for the suitability classifier.

Feature	Importance
Rainfall	0.2165
Humidity	0.2055
Potassium (K)	0.1731
Phosphorus (P)	0.1507
Nitrogen (N)	0.1081
Temperature	0.0781
pH	0.0606
Soil Moisture	0.0075

TABLE I. SUITABILITY MODEL FEATURE IMPORTANCE

Rainfall and humidity dominate, jointly accounting for over 42% of total importance, consistent with their strong agronomic relevance to crop viability. Notably, the soil moisture feature receives the lowest importance of all eight features (0.0075) — roughly 29 times lower than rainfall — which is consistent with it carrying little genuine predictive

signal, as would be expected of a feature not drawn from real, label-correlated measurements. This result is reported as supporting evidence for, rather than contradicting, the limitation disclosed in Section V-A.

X. IMPLEMENTATION

The system is implemented in Python using scikit-learn for both the suitability classifier and the per-crop price regressors, pandas for data handling, and Streamlit for the web interface. The interface exposes soil and weather input fields (with a random-sample generator for demonstration), a ranked results table showing suitability, forecasted price, monthly profit and risk level per candidate crop, a downloadable report combining input parameters and results, and a what-if scenario explorer with both quick preset buttons and free-text single-factor questions.

Crop	Suitability	Price (Rs./quintal)	Monthly Profit (Rs./ha)	Risk
jute	0.0900	5,183.1400	19,007.2500	Medium
rice	0.9030	3,107.6200	10,503.3300	Medium

XI. CASE STUDY: Suitability-Profit Divergence

To demonstrate the central contribution of this work, one soil/weather inputs were evaluated. In the first, with conditions favoring maize and cotton (N=110, P=45, K=20, temperature=24°C, humidity=65%, pH=6.8, rainfall=85mm), the suitability classifier ranked maize as substantially more suitable than cotton (72.7% vs. 27.0%). However, because cotton's forecasted price (₹7,119.60/quintal) and resulting monthly profit (₹51,793.96/ha) exceeded maize's (₹2,170.50/quintal; ₹33,114.98/ha), the system recommended cotton — the less agronomically confident choice — over the more suitable maize.

This divergence is the direct, demonstrable evidence for the paper's central claim: a suitability-only system would have recommended maize without ever surfacing cotton as a superior economic choice, and a price-only system (as in the closest prior work [4]) would have ranked correctly by price but without confirming, as this system does, that the profit advantage holds after cultivation cost is subtracted and normalized by crop duration.

XII. DISCUSSION

The results show that agronomic suitability and economic profitability can produce different rankings, potentially leading to lower-profit choices when considered separately. Duration normalization (Section VI-C) affected crop rankings, as high-yield crops such as banana (approximately 425 quintal/ha) may dominate per-cycle profit comparisons despite moderate price changes. This is an intended feature of profit-based ranking. The what-if tool is more effective when crops have similar baseline profits or larger price changes are simulated.

The elevated MAPE for mango and pomegranate (Fig. 2) highlights the limitation of the short 13-month historical window. The chronological 80:20 split leaves approximately two months for testing per crop, making MAPE values indicative rather than statistically precise. This limitation is transparently communicated through the risk label mechanism (Section VI-D).

XIII. LIMITATIONS

Several limitations remain in the current prototype. First, the system can recommend only from the 22 crops included in the dataset. Second, soil and weather inputs are treated as seasonal values rather than continuous real-time measurements. Third, profit estimation uses reference yield and cultivation-cost values instead of dynamically predicting yield. Fourth, price forecasting may be less reliable for crops with limited historical observations. Finally, the what-if tool currently supports only simple rule-based queries.

XIV. FUTURE WORK

Future work will focus on expanding the system to a larger number of crops and more comprehensive datasets. Real historical price data can be collected over longer periods to improve the reliability of crop-specific price forecasting. A yield-prediction model can also be developed to estimate crop yield based on soil and weather conditions rather than using fixed reference values. Further improvements can include integrating live weather and soil data, improving the what-if analysis, and evaluating the system across different regions and growing seasons.

XV. CONCLUSION

This paper presented a four-stage crop recommendation system combining a Random Forest suitability classifier, per-crop price forecasting models trained in part on real Government of India Agmarknet data, a duration-normalized profit engine, a forecast-reliability risk indicator, and an interactive what-if tool. A case study demonstrated the system recommending a less agronomically suitable but more profitable crop (cotton over maize), showing that suitability-only and price-only recommendation strategies can diverge from the profit-maximizing outcome.

The work extends the closest prior hybrid suitability-price system by computing genuine, duration-normalized profit rather than ranking by price alone, while also surfacing forecast uncertainty. The data and parameter limitations of the prototype are disclosed explicitly, consistent with its scope as a prototype system rather than a production deployment claim.

REFERENCES

- [1] M. K. Senapaty, A. Ray, and N. Padhy, "A Decision Support System for Crop Recommendation Using Machine Learning Classification Algorithms," *Agriculture*, vol. 14, no. 8, Art. no. 1256, 2024, doi: 10.3390/agriculture14081256.
- [2] S. R. Palle and S. A. Raut, "Crops Recommendation System Model Using Weather Attributes, Soil Properties, and Crops Prices," in *Sentiment Analysis and Deep Learning*, AISC, vol. 1432, Springer, 2023, pp. 323–338, doi: 10.1007/978-981-19-5443-6_24.
- [3] J. Daniel, R. Shyamala, R. Pugalenth, and P. Mohan Kumar, "RANC-CROP: Recommendation Attributed to Soil Nutrients and Stock Analysis Using Machine Learning," *IETE Journal of Research*, 2022, doi: 10.1080/03772063.2022.2060868.
- [4] N. M. Sindhur, C. Pavithra, and N. Muchikel, "A Hybrid Machine Learning Framework for Optimizing Crop Selection via Agronomic and Economic Forecasting," *arXiv preprint arXiv:2507.08832*, 2025.
- [5] S. Sam and S. M. D'Abreo, "Crop Recommendation with Machine Learning: Leveraging Environmental and Economic Factors for Optimal Crop Selection," *arXiv preprint arXiv:2505.21201*, 2025, doi: 10.48550/arXiv.2505.21201.
- [6] Y. Mahale, N. Khan, K. Kulkarni, S. A. Wagle, P. Pareek, K. Kotecha, T. Choudhury, and A. Sharma, "Crop Recommendation and Forecasting System for Maharashtra Using Machine Learning with LSTM: A Novel Expectation-Maximization Technique," *Discover Sustainability*, vol. 5, Art. no. 134, 2024, doi: 10.1007/s43621-024-00292-5.
- [7] P. Mahajan, A. A. Muley, and M. R. Fegade, "Forecasting Agricultural Commodity Prices Using Machine Learning Models," *Discover Informatics*, vol. 1, Art. no. 15, 2026, doi: 10.1007/s44564-026-00015-0.
- [8] M. Y. Shams, S. A. Gamel, and F. M. Talaat, "Enhancing Crop Recommendation Systems with Explainable Artificial Intelligence: A

Study on Agricultural Decision-Making,” *Neural Computing and Applications*, vol. 36, no. 11, pp. 5695–5714, 2024, doi: 10.1007/s00521-023-09391-2.

[9] Ö. Turgut, I. Kök, and S. Özdemir, “AgroXAI: Explainable AI-Driven Crop Recommendation System for Agriculture 4.0,” in *Proceedings of the 2024 IEEE International Conference on Big Data (BigData 2024)*, pp. 7208–7217, 2024, doi: 10.1109/BigData62323.2024.10825771.