KisanSevak: Enhancing Crop Management and Market Forecasting through Machine Learning

Yugchhaya Galphat *Department of Computer Engineering Vivekanand Education Society's Institute of Technology Chembur Mumbai 400074*

Shreerang Vaidya *Department of Computer Engineering Vivekanand Education Society's Institute of Technology Chembur Mumbai 400074* 2022.shreerang.vaidya@ves.ac.in

Jatin Navani *Department of Computer Engineering Vivekanand Education Society's Institute of Technology Chembur Mumbai 400074*

Samarth Nilkanth *Department of Computer Engineering Vivekanand Education Society's Institute of Technology Chembur Mumbai 400074*

*Abstract—***In today's digital era, despite technological advancements, farmers rely heavily on traditional farming methods. While these age-old practices hold cultural significance, there are inherent challenges that need addressing, including issues tied to choosing optimal cultivation strategies, accurate prediction of yield, and navigating market dynamics. This paper presents an innovative application of machine learning, transforming farming practices. This research explores predictive modeling for crop production based on soil and environmental conditions. It utilizes Support Vector Machines (SVM) for precise crop classification, Random Forest for accurate regression for yield prediction, and Long Short-Term Memory (LSTM) networks for nuanced price forecasting. The aim is sustainable agricultural intensification, focusing on factors like nitrogen, phosphorus, and potassium values to assess soil health and predict future harvests. Through the amalgamation of machine learning models and data-driven approaches, the research strives to bring about a fundamental shift in traditional agricultural practices, paving the way for widespread adoption of these prediction models across India.**

Keywords—Agriculture, Machine Learning, Crop Recommendation, Yield Prediction, Price Forecasting, Support Vector Machine, Random Forest, Long Short-Term Memory

I. INTRODUCTION

According to the Indian Economic Survey for the year 2020-21[19], the country globally stands at the position of the world's second-largest country for farm outputs, with agriculture employing over 54.6% of the Indian workforce with an 18.3% contribution to the country's economy by the year 2022-23[16]. Despite these figures, Indian farmers consistently endure significant hardships, whether in terms of economic or social growth. A considerable portion of India's economic development hinges on effective agricultural planning. According to a Harvard review[17], food demand may rise from 59% to 98% by 2050, but current crop yield may not significantly meet those demands. Therefore, it is crucial to apprise farmers of the anticipated production from their fields and also help them comprehend the impact of climate change on their lands, facilitating better monitoring of crop growth[19].

Local traders, often acting as monopolies and creditors, possess substantial power over farmers, offering

Anjali Thakrani *Department of Computer Engineering Vivekanand Education Society's Institute of Technology Chembur Mumbai 400074*

non-negotiable profit margins. Although initiatives such as Farmer's Portal^[18] exist to aid farmers, they lack modernity and personalization leading to issues like unorganized content, site malfunctioning, scattered data, and lack of personalization. To overcome these issues a model solution can be implemented to exclusively predict the crop yield and crop prices efficiently. A diverse machine learning (ML) technique can be implemented taking into account parameters such as live weather conditions, soil health, etc. These proposed systems not only help to improve the prediction of the net yield of crops but also enhance farming techniques to minimize crop loss.

ML tools, particularly for crop recommendation, also aid in yield prediction at different scales, from local to national level. By analyzing historical data such as soil composition, weather patterns, and crop performance models can recommend suitable crops for specific farms, optimizing yield and profit. Various linear and non-linear regression models were compared, in addition to time-series analysis and forecast help in suggesting and selecting crops to cultivate based on price predicted using Machine Learning.

By aggregating data on market trends and factors influencing prices, a sophisticated neural network can offer valuable insights. The key contributions encompass three vital aspects: first, the development of a localized, indigenous solution that seamlessly applies machine learning. This entails formulating a farmer-friendly system featuring an easily navigable interface, accurate predictions, personalized and secure data analysis, and an integrated feedback loop to ensure ongoing model accuracy. Secondly, an indigenous solution involves the integration of data from local markets, a deep understanding of supply chain dynamics, and the incorporation of cultural practices influencing pricing. This approach ensures that predictions are not only more accurate but also highly relevant to the specific community in question. For broader adoption, particular attention should be paid to designing the user interface, taking into account local languages and literacy levels. This user-centric design ensures that the technology is accessible and user-friendly, facilitating widespread use and understanding within the target community.

II. LITERATURE SURVEY

The grassroots of the proposed solution was implemented on a thorough analysis of existing literature. Noteworthy contributions by A. Sharma et al. [14], explore the quantifiable aspects of agriculture affecting plant growth and relevant machine learning models applicable to precision agriculture. S. Qazi et al. [15] talk about the application of IoT in agriculture, smart irrigation techniques like hydroponics, use of deep learning in agriculture, the problems it can address and solve, and future trends for smart farming. Additionally, a comprehensive review of existing literature pertaining to price forecasting algorithms in the context of agricultural commodity prices was conducted.

In the research conducted by Zhang et al. the exploration of Long Short-Term Memory (LSTM) networks and Temporal Convolutional Networks (TCN) involved testing for forecasting soybean prices in both univariate and multivariate settings [4]. Kurumatani et al. introduced two approaches for commodity price prediction: Time-alignment of Time Point Forecast (TATP), which generates individual values for each point in time, and Direct Future Time Series Forecast (DFTS), which produces a sequential time sequence [5]. The study considers a range of factors including past price history, sales, supply amount, consumer preference, and weather parameters for forecasting. However, the emphasis of this research lies on market price data. In contrast, Ladhar et al. focused primarily on feature extraction and evaluated different models, including simple Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Extreme Learning Machines (ELMs) [1]. The study incorporated parameters influencing soybean price prediction such as historical prices and arrivals, surrogate market prices, as well as news-derived event features and day of the year information. The evaluation of models like Long Short-Term Memory (LSTM) and Temporal Convolutional Networks (TCN), with TCN utilizing surrogate markets, performed best for short-term forecasts (of 30 days).

Moreover, an in-depth and thorough exploration of pertinent literature concerning crop recommendation and yield prediction was diligently carried out. The detailed descriptions of these studies are presented in Table I.

TABLE I. SUMMARY OF LITERATURE RELATED TO CROP RECOMMENDATION AND YIELD PREDICTION

III. PROPOSED SYSTEM

Despite significant research efforts in the agricultural domain, there remain critical gaps that hinder the development of comprehensive solutions for farmers. This paper not only identifies key areas necessitating further research but also proposes a unified system designed to enhance the effectiveness of agricultural practices. The suggested solution takes the form of an application empowering farmers to make informed decisions on crop cultivation. It helps the farmer decide the crop to be cultivated, predicts the yield of a particular crop sown, as well as a module that forecasts market prices.

A. Collection of Dataset and Preprocessing

Data like the ratio of nitrogen, phosphorus, and potassium (N: P: K ratio), the annual rainfall, area of a given field, year of cultivation of a given field, crop name, average temperature, pesticide usage in tonnes, season, etc. were obtained, along with time series data for price prediction, which was scraped from the government website AgMarknet[20]. Initially, missing values from historical data and temporal gaps in forecasting data were meticulously filled in. Following this, features about seasonal data were synthesized through an in-depth analysis of relevant regional data. To enhance the accuracy of the results, redundant features were pruned based on their correlation with other variables, such as the cultivation year and field area. The

refined list of features selected for the three models is outlined below:

- Crop recommendation: After getting relevant data from the farmer, the farmer gets to know what crop is most suitable for the parameters of their farmland and location. Currently, the recommendation system can recommend one out of a total of 22 crops.
- Yield prediction: The yield of the planted crop can be predicted using this module.
- Price forecasting: Market prices of the next day are predicted daily and given to the farmer.

Fig. 1 depicts a block diagram of the proposed system.

Fig. 1. Architecture of the proposed system.

B. Crop recommendation

A simple classification model was made for crop recommendation. For the same, Support Vector Machine (SVM), K-Neighbours, and Random Forest were compared after training and evaluation of the models. Out of the three, Support Vector Machine was finally selected, giving the best results.

Support Vector Machines (SVMs) prove highly effective in supervised learning, particularly for classification tasks. They excel in identifying the optimal hyperplane within a multi-dimensional space, acting as a boundary that effectively segregates data points of different classes. The model's robustness is enhanced by a wider margin between this hyperplane and the support vectors, enabling superior generalization to new data. For example in Fig. 2, a hyperplane (in this case, a 2 dimensional boundary), classifies data on 2 arbitrary features.

In our particular case, the model follows a One-vs-Rest strategy to predict the appropriate crop. In this strategy, a

binary classifier is trained for each class against the rest. After that, the classifier with the highest score is selected as the predicted class. Given below is a formula for the same:

$$
f_i(x) = sign(w \cdot x + b) \tag{1}
$$

Here, w_i is the weight for a particular binary classifier and *b* is the bias. The output is either -1 or 1. The function of the SVM is to find the optimal weight and bias that maximizes its margin while ensuring that all data points are correctly classified. SVMs are particularly advantageous when dealing with features exhibiting clear boundaries, as illustrated in Fig. 3. The figure demonstrates the clustering of crop data by plotting N: P: K ratio values against each other. With a semblance of discernible data clusters, the suitability of Support SVM becomes evident, presenting a well-founded choice for our analytical approach.

Fig. 2. Example SVM hyperplane representation.

Fig. 3. Correlation of Nitrogen, Phosphorus and Potassium levels.

C. Yield prediction

After a thorough evaluation of various regression models, including Random Forest, Support Vector Machine, and Ridge regression, the Random Forest model emerged as the optimal choice, exhibiting superior performance on the test data.

Random Forest regression, a subset of the ensemble learning paradigm, harnesses the power of multiple decision trees to achieve accurate and stable regression predictions. Each decision tree in the ensemble employs a tree-like model, mapping out decisions and their potential consequences through a sequence of conditional branching. An example of a single decision tree is shown in Fig. 4. While individual decision trees can handle non-linear relationships, they are susceptible to overfitting and high variance. The ensemble approach effectively mitigates these issues, as each tree independently processes a subset of the data. The final prediction often results from aggregating the individual tree outputs, typically calculated as the mean.

Fig. 4.Sample Decision Tree.

Random Forest proves pivotal in our approach, excelling with its adeptness at deciphering intricate non-linear relationships within complex datasets. This capability positions it as a powerful tool for extracting nuanced insights crucial for our model.

D. Price forecasting

In the pursuit of accurate price forecasting, an initial exploration involved studying and decomposing the dataset to find its seasonality. As illustrated in the graph of daily price deviation in Fig. 5., the data exhibits a discernible weekly seasonality, indicating recurrent patterns in trends every 7 days. Subsequently, a time series forecasting model was developed, employing the Long Short-Term Memory (LSTM) architecture. This neural network was employed with a historical data input span of the preceding 56 days.

LSTM networks are a specialized form of recurrent neural networks designed to capture extended dependencies in sequential data while addressing the vanishing gradient problem. Memory cells are featured with input, output, and forget gates, allowing for selective retention and forgetting of past information based on current inputs. This enables effective learning and prediction in sequential tasks such as speech recognition and time series prediction. LSTM networks excel in modeling long-range dependencies and temporal dynamics, making them valuable tools in sequential data analysis and prediction. They're particularly practical when dealing with time series data due to their effective handling of temporal patterns.

Fig. 5. Graph exploring weekly seasonality in temporal data.

Fig. 6. Design of an LSTM cell.

Fig. 6 depicts a single unit (i.e. cell) of the network. Listed below are the relevant components of the cell:

- Cell State (C) : It represents the memory of the LSTM cell, crucial for storing information across time steps.
- Forget Gate (f): This gate controls how much information must be retained from the previous cell state (C_{t-1}) .
- Input Gate (i): This gate determines which information from the current input (X_t) is stored in the cell state.
- Output Gate (o): This gate controls the information from the current cell state (C_t) that forms the output.

The forget gate works on the following formula:

$$
f(t) = \sigma(W_f \cdot h_{t-1} \cdot X_t + b_f) \tag{2}
$$

Here σ is the activation function with range (0, 1), W_f is the weight matrix of the forget gate, h_{t-1} which is the previous hidden state, X_i being the current state, and b_f is the bias vector of the forget gate. Values closer to 1 in *f(t)* indicate retaining more information, while values closer to 0 signify forgetting more. The hidden state carries the processed information and is used by subsequent LSTM cells or serves as the final output depending on the application. By processing information through these gates and the cell state with their associated weight matrices and biases, LSTMs demonstrate an effective ability to capture prolonged dependencies in sequential data.

The culmination of this research effort materialized in the development of a comprehensive application featuring distinct modules, encompassing a sophisticated crop recommendation page, an accurate yield prediction page, an insightful price forecasting page, and a resilient overall architecture, thereby establishing a robust and multifaceted tool for agricultural decision-making. This is reflected in the figures 7, 8, 9, and 10. This app ascertains a personalized experience and it incorporates regional data.

Fig. 7. GUI Screenshot of landing page.

Fig. 8. GUI Screenshot of crop recommendation page.

Fig. 9.GUI Screenshot of yield prediction page.

Fig. 10. GUI Screenshot of market price forecasting page.

A. Evaluation and Results

For the crop recommendation model, precision, recall, and F1-score were used as evaluation measures.

$$
Accuracy = \frac{TP + TN}{TP + TN + FP + FN}
$$
 (3)

$$
Precision = \frac{TP}{TP + FP}
$$
 (4)

$$
Recall = \frac{TP}{TP + FN} \tag{5}
$$

$$
F1 = \frac{2 \cdot Precision\cdot Recall}{Precision + Recall} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}
$$
 (6)

For the evaluation of the yield prediction model as well as the price forecasting model, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were calculated.

$$
MAE = \sum_{i=1}^{D} |x_i - y_i|
$$
 (7)

RMSE =
$$
\sqrt{\frac{1}{n} \sum_{i=1}^{D} (x_i - y_i)^2}
$$
 (8)

After the comparison, the Support Vector Machine was found to be the most accurate for crop recommendation, based on the precision, recall, and F1 score. For yield prediction, Random Forest gave the best results overall. Tables II and III contain the results obtained on training the various models that were tested.

Model	MAE	RMSE
Random Forest	0.00367	0.0243
Support Vector Machine	6.71	11.4
Ridge regression	7.16	9.15

TABLE III. RESULTS OF REGRESSION MODELS

IV. CONCLUSION AND FUTURE WORK

Indian farmers have always encountered the problems of price fluctuations, climate hazards, chronic indebtedness, and much more. Also, these problems become more vulnerable for small landholding farmers or farmers who are unaware of the current market prices or the best crop to yield at a particular time. The crop recommendation model relied on a Support Vector Machine (SVM), proving effective in providing accurate results. For yield prediction, the Random Forest regression model demonstrated high accuracy. LSTM (Long Short-Term Memory) emerged as the optimal choice for predicting crop prices and fluctuations. The presented solution helps the farmer anticipate these problems by not only recommending a suitable crop for the land but also helping in predicting the yield with an awareness of the ongoing and past prices of the crop. An indigenous solution is obtained that lucidly applies machine learning for the farmers of India by building a farmer-friendly system and a step to solve problems at an earlier stage.

The proposed agricultural system could significantly benefit from real-time web scraping and integration, allowing it to dynamically gather the latest data on climate, market prices, and relevant agricultural information. By incorporating this real-time data, farmers could receive up-to-the-minute insights, enabling more informed decision-making. Additionally, implementing a robust feedback mechanism within the system allows farmers to provide valuable insights based on their experiences.

REFERENCES

- [1] Ronak Ladhar, Sourav Sharma, Srikant Tangirala, Nishant Gupta, Abdul Azeem, Arjav Jain, Bhuvan Chand Katakam, Aditya Bommakanti, C. Sankaraiah, Hari Prasad Piridi, Kaushalendra Yadav, Kumra Vittalrao, Matiur Rahman, Rashul Chutani, Rishi Shah, Rohan Katepallewar, Dipanjan Chakraborty, and Aaditeshwar Seth. 2023. AI-based Market Intelligence Systems for Farmer Collectives: A Case Study from India. ACM J. Comput. Sustain. Soc. 1, 1, Article 6 (September 2023), 32 pages. https://doi.org/10.1145/3609262
- [2] Mupangwa, W., Chipindu, L., Nyagumbo, I. et al. Evaluating machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa. SN Appl. Sci. 2, 952 (2020). https://doi.org/10.1007/s42452-020-2711-6
- [3] Y. J. N. Kumar, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi, "Supervised Machine Learning Approach for Crop Yield Prediction in Agriculture Sector," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 736-741, doi: 10.1109/ICCES48766.2020.9137868
- [4] D. Zhang, S. Chen, L. Liwen and Q. Xia, "Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons," in IEEE Access, vol. 8, pp. 28197- 28209, 2020, doi: 10.1109/ACCESS.2020.2971591.
- [5] Kurumatani, K. Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method. SN Appl. Sci. 2, 1434(2020). https://doi.org/10.1007/s42452-020-03225-9
- [6] S. Salankar, A. Salankar, A. Sune, P. Suryavansh and H. Kumar, "Crop Suggestion using Data Mining Approaches," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-5, doi: 10.1109/ICCCNT51525.2021.9579
- [7] R. Varun Prakash, M. Mohamed Abrith and S. Pandiyarajan, "Machine Learning Based Crop Suggestion System," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2022, pp. 1355-1359, doi: 10.1109/ICICCS53718.2022.9788314.
- [8] P. A, S. Chakraborty, A. Kumar and O. R. Pooniwala, "Intelligent Crop Recommendation System using Machine Learning," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 843-848, doi: 10.1109/ICCMC51019.2021.9418375.
- [9] D. Modi, A. V. Sutagundar, V. Yalavigi and A. Aravatagimath, "Crop Recommendation Using Machine Learning Algorithm," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-5, doi: 10.1109/ISCON52037.2021.9702392.
- [10] Z. Doshi, S. Nadkarni, R. Agrawal and N. Shah, "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697349.
- [11] R. J, V. K. G. Kalaiselvi, A. Sheela, D. S. D and J. G, "Crop Yield Prediction Using Machine Learning Algorithm," 2021 4th International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2021, pp. 611-616, doi: 10.1109/ICCCT53315.2021.9711853.
- [12] D. Elavarasan and P. M. D. Vincent, "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications," in IEEE Access, vol. 8, pp. 86886-86901, 2020, doi: 10.1109/ACCESS.2020.2992480.
- [13] M. Singh Boori, K. Choudhary, R. Paringer, A. Kupriyanov and Y. Kim, "Wheat Yield Estimation and Prediction Via Machine Learning," 2023 IX International Conference on Information Technology and Nanotechnology (ITNT), Samara, Russian Federation, 2023, pp. 1-4, doi: 10.1109/ITNT57377.2023.10139117
- [14] A. Sharma, A. Jain, P. Gupta and V. Chowdary, "Machine Learning Applications for Precision Agriculture: A Comprehensive Review," in IEEE Access, vol. 9, pp. 4843-4873, 2021, doi: 10.1109/ACCESS.2020.3048415.
- [15] S. Qazi, B. A. Khawaja and Q. U. Farooq, "IoT-Equipped and AI-Enabled Next Generation Smart Agriculture: A Critical Review, Current Challenges and Future Trends," in IEEE Access, vol. 10, pp. 21219-21235, 2022, doi: 10.1109/ACCESS.2022.3152544.
[16] Press Information Bureau.
- Information Bureau. (n.d.). <https://www.pib.gov.in/PressReleasePage.aspx?PRID=1741942> (accessed March. 4, 2024)
- [17] Elferink, M. (2016, April 7). Global demand for food is rising. Can we meet it? Harvard Business Review [https://hbr.org/2016/04/global-demand-for-food-is-rising-can-we-meet](https://hbr.org/2016/04/global-demand-for-food-is-rising-can-we-meet-it) $-it$ (accessed Feb. 12, 2024)
- [18] Farmer Portal : Home page. (n.d.). <https://www.farmer.gov.in/> (accessed Jan. 30, 2024)
- [19] Ministry of Finance, Government of India (n.d.). [https://www.indiabudget.gov.in/budget2021-22/economicsurvey/doc/e](https://www.indiabudget.gov.in/budget2021-22/economicsurvey/doc/echapter_vol2.pdf) [chapter_vol2.pdf](https://www.indiabudget.gov.in/budget2021-22/economicsurvey/doc/echapter_vol2.pdf) (accessed Dec. 26, 2023)
- [20] Agriculture Marketing. (n.d.). <https://agmarknet.gov.in/> (accessed Feb. 24, 2024)