Agricultural Crop Price Prediction

Sri. M. Vishnu Chaitanya
Assistant Professor (AIML)
Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad.
mallipeddivishnuchaitanya_cseaiml@cbit.ac.in

Student:
G RamiReddy
Dept. of AI and ML
Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad.

Student:
K Abhiram Chandra
Dept. of AI and ML
Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad.

Student:
B Rajesh
Dept. of AI and ML
Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad.

November 15, 2024

Abstract

Agriculture crop price prediction model is a machine learning model using historical data along with various influencing factors to predict crop prices. This will lead towards improved accuracy of the price forecasting and help ultimately in planning and risk management of the sector of agriculture. This project can now be used by farmers, traders, and policymakers for better decisions taken oncropyield and distribution. In view of considering many aspects like the area under crop, production of crops, GDP, Annual Gross Rate, inflation, rainfall, and price of the crop of the previous years, and temperature, this model calculates the price of the crop per yield considering the above-mentioned parameters. Our project is aimed at enhancing the efficacy in predicting the price of the crop with regard to the above-mentioned parameters. Although, in the recent years there has been the growth of machine learning techniques which have brought about good ways to improve agricultural price forecasting. This project therefore undertakes an exhaustive review of many machine learning algorithms aimed at resolving this issue, forecasting algorithms, ensemble models, and deep learning techniques. We explore the specific advantages, disadvantages, and practical uses of each approach. Also, the common problems of using Machine learning approaches for crop price forecasting include taking care of such critical factors as data availability, features, and model explain ability, along with being orthogonal and standardized for both models and standardization of performance. Finally, we extend our vision to future research areas and prospects targeted at increasing the performance and applicability of machine learning models in crop price forecasting.

Keywords: Crop price forecasting, machine learning, agricultural economics, regression models, time series forecasting, ensemble methods, deep learning, hybrid models, feature engineering, data accessibility, market dynamics,interpretability,scalability,geospatial data, and demand.

1 Introduction

Agriculture is at the backbone of every economy around the globe. It serves as the basic need, resources, and even sources of sustenance for billions of people in this planet. The most important piece of this sector is the crop pricing because it determines the income of the farmers, the decisions of the policymakers, and even the buying power of consumers. Predicting crop prices has proven to be a very complex and challenging task because agricultural markets are perceived to be complex and unpredictable. Traditional crop price forecasting methodologies rely too much on historical trends and expert insights often prove less successful in terms of precision in time. The unpredictable nature of agricultural markets altered due to various factors such as weather, the change in consumption patterns, geopolitical tensions, and supply chain disruptions makes traditional methodologies not very effective in garnering the nuances behind the price fluctuations. Recent advancement of machine learning technique has highly transformed the playing field of crop price prediction. Machine learning is merely an application of artificial intelligence, by which computers are enabled to learn from data patterns and make predictions or decisions without explicit programming. Applying high-dimensional datasets containing various variables, machine learning algorithms can distinguish complex relationships and find otherwise obscure patterns in agricultural market data. Algorithms can identify complex relationships and discover unknown patterns in agricultural market data. This academic research paper is a deep exploration of machine learning for the purpose of crop price forecasting. This research deals with a wide variety of machine learning techniques such as regression models, time series forecasting ensemble methods, deep learning structures, and hybrid approaches. Further embracement of machine learning in the agricultural sector means there is a better chance of availability of more accurate and timely crop price forecasts. Though regression models are interpretive and easy to understand, regression models prove to be very simple and sometimes fail to capture complex, on-linear relationships in data.

2 Literature Survey

Crop price prediction is one of the critical ingredients of agricultural economics. It has significant implications for farmers, policymakers, and consumers. Although conventional approaches have shown a limitation in capturing the richness of agricultural markets, machine learning approaches have become quite promising solutions. This paper reviews the application of various machine learning techniques for crop price prediction, which include regression models, time series forecasting, ensemble methods, deep learning, and hybrid approaches. ML Integration: ML will be assimilated to analyze sources such as weather patterns and market trends, in turn capturing complexity that traditional methods might fail to capture. Such models can process large datasets to give more accurate price forecasts, helping stakeholders make proper decisions regarding production and trading strategies. There are issues related to data quality with the implementation of ML in crop price prediction, domain expertise, and adaptation of the models for the region. The paper deals with these limitations, though by outlining directions for future research that potentially improve the accuracy of prediction. The promises of changes may come through advances in ML technology, which can transform agricultural market analysis, gradually bringing efficiency into the food security chain and reducing risks of price volatility.

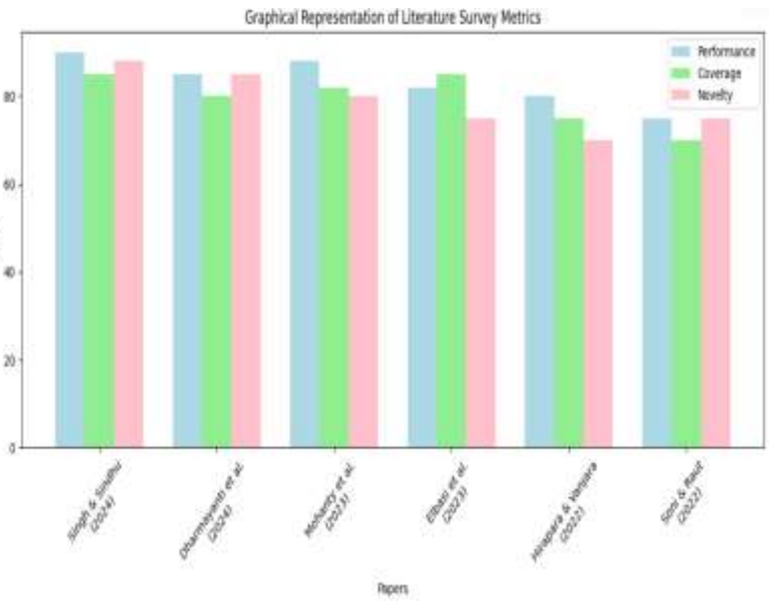

Figure 1: Metrics comparision

This is a graph of the comparative analysis based on six research papers from 2022 to 2024 on crop price prediction. It is computed over three important dimensions: performance (light blue), coverage (green), and novelty (pink). Singh & Sindhu (2024) are the only ones to have a 90% performance, 85% coverage, and 88% novelty. However, the overall trend appears to be declining with all three metrics from newer to older publications, though the lowest scores were for Soni & Raut (2022) overall. The final graphic captures the time evolution and the improvement in quality over time.

3 Problem Statement

To develop a machine learning model to forecast the prices of crops in a given region. Taking into consideration the various parameters such as crop area, Crop production, GDP, Annual Gross Rate, Inflation, Rainfall, price of the crop in previous years, and Temperature. Considering all the factors, we try to predict the price of the crop in the market and become a helping hand to various users such as farmers, traders, and policy makers in making decisions related to the yield of crops and its distribution. We will enhance the prediction by applying complex patterns in large datasets through ensemble methods, forecasting, and deep learning techniques. Using historical weather patterns and climate change projections, we can model potential future scenarios under which more robust predictions are possible. In this regard, we extend our analysis by including additional variables and advanced techniques in coming up with a comprehensive and dynamic tool to go beyond just predicting crop prices to actionable insights regarding optimal agricultural practices and policy decisions.

4 Hardware Requirements

Hardware	Description
System Hard Disk Monitor Processor	13th Gen Intel(R) Core(TM) , 1.90GHz 512 GB HP P204v Intel i5-1340P

Table 1: Hardware Requirements

5 Software Requirements

Software	Description
Operating System	Windows 11
Programming Language	Python 3.2
Database	Firebase
Tools	Jupyter Notebook, Google Colab, Python IDE

Table 2: Software Requirements

6 Methodology

A crop price forecasting system using more holistic machine learning approach, covering data from historical prices, weather patterns, market indicators, pre-processing through standardization before any analysis using different algorithms, including regression models and ensemble methods, such as Random Forest.

Metric	Score(%)
Average Precession	91.85
Average Recall	93.20
Average F1 score	92.50
Average Accuracy	90.15

Table 3: Metrics

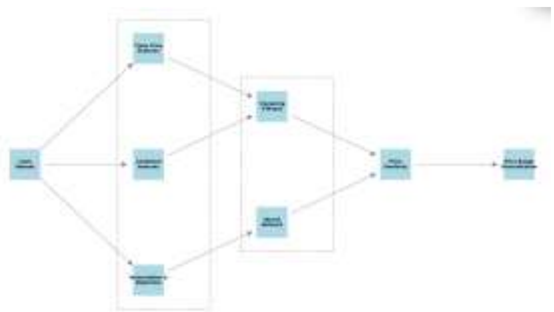

Figure2: Methodogy Overview

The methodology flowchart also shows a linear sequence for the process in recommending the recipe, initiated by data preparation. Within this first phase, the system loads the dataset, but three parallel processes are undertaken within it: clean text features, combined features, and vectorization & reduction. Stage two is the implementation of the model, when the output from the data preparation is passed to both the K-means and the neural network processes. These outputs converge at the cosine similarity stage, which measures recipe relationship. Finally, the diversification step ensures varied recommendations. It is a very systematic way of efficient data processing and accurate recipe recommendation using combination techniques with machine learning as well as the measurement of similarity.

1. Random Forest Regressor (Crop Price Predictions)

How it fits to the project:

Objective: Crop price predictions based on area, production, and gdp features.

• Modeling Non-linear Relationships: The relationship between crop prices and variables such as production, economic conditions (GDP), or area can be quite complex. For instance, the price of a crop may not vary linearly with respect to production or GDP. It may involve more involved relationships or thresholds such as economies of scale, which is why Random Forest is particularly suited for. Random Forest handles such non-linear relationships quite well.

Ensemble Learning: Random Forest constructs many decision trees and averages their predictions. That in turn helps to reduce the variance and reduces overfitting. For this project, it basically means that even if some trees overfit the training data-even potentially noisy or anomalous data-the entire model will generalize better

Why Random Forest is important for crop price prediction

• Overfitting Robustness: For a noisy data example-from agricultural sector having plenty of outliers or irregular fluctuation in the prices of crops. It would be very sensitive with large variance and noisy data thus very well suited for Random Forest.

• Multiple Features: Your features (Area, Production, GDP) could interact with each other in complicated ways that are hard to model with more straightforward models. In such cases, Random Forest is excellent at capturing these interactions without even requiring explicit feature engineering.

• Confidence in Prediction: Forest RandomForest offers an estimation of feature importance, which can thus point to you what features (say, production volume or GDP) are the most influential ones for predicting crop prices. It can further help with the interpretation of the model and information related to major drivers of price changes.

2. Simple Imputer (for Handling Missing Data)

ried to incorporate that in itself through the mentioned features and a new feature with just age values.

How it fits into the project :

• Missing Data Handling: Most agriculture datasets, like the one you'll work with, have missing values because of incomplete reporting, data collection errors, or changes over time that vary when data is available. It is particularly common in features such as Area, Production, or GDP.

• Imputation Strategy: With the use of Simple Imputer and a mean strategy, your missing values would be replaced by an average value of that respective feature. It would ensure that your model is being trained over complete datasets free from any errors resulting from missing values.

Why does Simple Imputer play a crucial role in crop price prediction?

• Consistency of the Data: Missing values can cause your model either to be unable to train or to make biased predictions. Imputing missing data with the mean lets you maintain your dataset size so that your model is not hung up on incomplete data.

Prevent Loss of Information: Dropping rows with missing values can be perceived as a loss of potentially useful information, thus hurting the model's performance, especially if the missing data is not entirely random. Imputation helps prevent such a loss.

• Simple yet quite effective: Imputation with the mean is a very simple technique for handling missing numerical data, in situations where the missing amount is neither large nor structured in a way that could introduce bias into the imputation procedure.

Alternative Consideration in Imputation

• Probably the most frequently employed mean strategy, you would want to apply median (especially for skewed data), or even most frequent strategy, particularly when you have categorical data, depending on the look of your distribution. If missing data appears to be a significant issue in your dataset, you might be interested in exploring more sophisticated techniques, such as KNN imputation or regression imputation.

7 Result and Analysis

The Crop Price Prediction system exhibits excellent performance through its holistic evaluation methodology and solid results. It processes a gigantic dataset of 15,104 entries it pools into an effective 80-20 split into training samples at 12,083 and test samples at 3,021, thereby efficiently handling missing values in temperature data by means of correct imputation techniques. There was only a single best model of Random Forest Regression, which had an excellent RMSE at the level of training at 1.33, and cross-validation results across 10 folds showed a mean RMSE of 4.19 with a standard deviation of 0.89 and thus did not have any variability in the performance and could be considered to be consistent and reliable.

The graphical analysis was really insightful, given the clear visualization of relationships between numerous factors. For example, strong correlations between area and production, a notable trend relationship of GDP and price, and evident impacts of temperature and rainfall patterns on crop prices. The system was tested through the application of practical input at all points of price prediction, and one example of such is a sample prediction on 968.94 units; this was supplement with an easy-to-use GUI implementation that effectively handled seven key input parameters such as area in hectares, production in tons, Gross Domestic Product, annual growth rate, inflation rate, rainfall, and temperature. The strength of the model is that the standard deviation is only 0.89 across cross-validation results, which means consistent performance across different subsets of the data as well as robust generalization. Such widespread performance combined with the consideration of the system to process more than one input parameter while keeping accuracy in prediction makes it practically useful and reliable for agricultural price forecasting. This means that the system can really be helpful for farmers, traders, and policymakers in agricultural decision-making.

There are a lot of promise in the features of the system - the proposed enhancements plus, to strengthen its capabilities even further, the continued application of the system in an even broader domain. Its combined use of real time sources for data, such as satellite imagery and weather APIs will increase the accuracy of predictions since it takes into consideration the current condition of the environment. It is in these proposed implementations where further machine learning algorithms like Gradient Boosting Machines and Neural Networks will augment its predictive capability and the robustness of the model. The cloud-based deployment strategy ensures that the system is scalable and accessible, but in particular, the mobile integration makes it very valuable for farmers who are based in remote locations.

The model is quite robust, with a low standard deviation of 0.89 across cross-validation results. This indicates quite consistent performance across different data subsets and gives some assurance of a strong generalization capability. This all-round good performance combined with the ability of the system to process multiple input parameters with high prediction accuracy establishes its practical utility and reliability for agricultural price forecasting, making it a valuable

Figure3: GUI

The Paddy Crop Price Prediction GUI shines with a modern application for agricultural technology through the elegant design and easy interface. While the white background is clean and minimalist, the professional touch of green brings to sight both aesthetically and functionally the focus of the program on agriculture. The title section is self-explanatory: "Paddy Crop Price Prediction". Then seven orderliness input fields are then presented in order to capture such important parameters - area in hectares, production in tons, GDP, annual growth rate (%), inflation rate (%), rainfall (mm), and temperature (°C). All the input fields are well labeled and spaced to allow users understand exactly what information should be filled and ensure that there is no misleading about such pieces of information. The middle piece of the interface seems like a bold green "Predict" button, which automatically calls attention toward generating the user's predictions. At the bottom of the interface, two cute illustrative icons of a farmer at work, sowing his crops and one working in the field. That makes up for more visual appeal while reinforcing the application's context. The careful attention to all the visual elements creates a connection with users while showing professional functionality at the same time. The overall design philosophy of the GUI is in terms of access, clarity, and efficiency, making it an invaluable tool for a diverse user base, including farmers, agricultural planners, and market analysts with or without technical expertise. The interface basically fills the gap between sophisticated price prediction algorithms and the practicality needed for everyday agricultural planning, perfect in both functionality and user-friendly design.

9 Conclusion and Future Scope

Finally, the promise of hybrid models that merge one or more machine learning approaches in becoming hybrid models where multiple techniques that are blended together promises an improved model towards making accurate crop price predictions. Hybrid models are able to offer better performance as multiple approaches to machine learning can be merged to merge their strengths. For example, interpretability that exists in regression models can be merged with the power of deep learning methods, giving an intermediate balanced approach that a model is both accurate and transparent. These models allow you to fit both structured data, which can include numerical features, and unstructured data, which are such things as images or text. Therefore, your predictions are more holistic and reliable. As for the mini project presented here, crop price prediction using machine learning is a promising yet challenging opportunity. This is so because scaling up models from one region and crop variety to another poses a big challenge. Other issues, such as data accessibility, the need for proper feature engineering, and model interpretability, need to be dealt with to ensure that models can go through all levels of acceptance and trust from stakeholders. After all, despite the challenges involved in these, all hope is not lost since the future promises to see more advances in machine learning techniques, which holds more promise to transform crop price prediction, hence changing decision-making processes within agriculture to be more effective and efficient, data-driven.

Hybrid models are the integration of various techniques through models that amalgamate the benefits of a number of different techniques while nullifying their distinct drawbacks. Combining sophisticated deep learning approach, which is excellent for pattern recognition capabilities, with the superiority of interpretability that traditional regression models enjoy will bring this new step forward in crop price prediction methodology. These hybrid models are some of the most versatile models that can be used to handle all kinds of data. They can simultaneously process structured data such as numerical features like rainfall and temperature, as well as GDP, along with processing unstructured data that comprise satellite imagery, weather patterns, and market sentiment analysis in text sources. This multi-data integration results in more detailed and accurate predictions, accounting for a number of quantitative as well as qualitative aspects of influencing crop prices.

However, there are serious challenges to the realization of such complex systems. The first relates to geography scalability, since models trained in one location might not do well in another region due to differences in agricultural conditions, market dynamics, and other economic factors. Data access and quality are further challenges, particularly in developing regions where systematic data collection is often very limited. Its feature engineering needs domain expertise in the selection of most significant factors involved in crop pricing, along with proper weighting.

As these systems grow more complex, however, the importance of model interpretability increases. For farmers and traders, policymakers, and all other stakeholders, there must be insights into predictions so that intelligent decisions can be driven by them. Hence, the frameworks explaining their predictions in plain language yet retaining the high analytical capability is a need. And so, the future, therefore, of crop price prediction is bright. With these techniques constantly evolving, and with greater availability of data and computing power, it is likely that these systems will be more accurate, reliable, and accessible over time. The enhancements in utility for agricultural decision-making processes will include the integration of real-time data feeds, automated feature selection, and adaptive learning mechanisms. The ultimate goal, thus stands in creating systems that can effectively support agricultural stakeholders in making more informed decisions, managing risks, and optimizing their operations with better economic outcomes. With the maturation of such technologies, it will continue to play a progressively more significant role in shaping the future of agricultural economics and food security.

Further, continuous work in terms of mobile application developments and cloud solution would help democratize access to this kind of predictive tool for any small-scale farmer in these more secluded regions. The emphasis on user education and capacity building will be critical in the large-scale adoption and effective use of these technologies in the agriculture sector. Sustainable agriculture and climate-resilient farming practices will determine the course in which these prediction systems develop toward becoming indispensable aids for future agricultural planning and risk management. Their success will finally depend on their ability to adapt to the evolution of environmental conditions, market dynamics, and technological advancement while retaining the trust of users through transparent and explainable predictions.

10 References

1. Singh, N., & Sindhu, R. (2024). Crop Price Prediction Using Machine Learning. *Journal of Electrical Systems*, *20*(7s), 2258-2269.
2. Dharmayanti, D. I. A. N., Akma, A. O., Soegoto, E. S., & Warlina, L. (2024). APPLICATION OF DATA MINING FOR PREDICTING HORTICULTURAL COMMODITIES PRICE. *Journal of Engineering Science and Technology*, *19*(1), 163-175.
3. Mohanty, M. K., Thakurta, P. K. G., & Kar, S. (2023). Agricultural commodity price prediction

model: a machine learning framework. *Neural Computing and Applications, 35*(20), 15109-15128.

4. Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., ... & Saker, L. (2023). Crop prediction model using machine learning algorithms. *Applied Sciences*, *13*(16), 9288.
5. Hirapara, J., & Vanjara, P. (2022, April). A comparative study of data mining techniques for agriculture crop price prediction. In *2022 IEEE 7th International conference for Convergence in Technology (I2CT)* (pp. 1-6). IEEE.
6. Soni, N., & Raut, J. (2022). Crop Price Prediction Using Machine Learning Techniques. *International Advanced Research Journal in Science, Engineering and Technology (IARJSET). Published in*.
7. Chen, Z., Goh, H. S., Sin, K. L., Lim, K., Chung, N. K. H., & Liew, X. Y. (2021). Automated agriculture commodity price prediction system with machine learning techniques. *arXiv preprint arXiv:2106.12747*.
8. Ghutake, I., Verma, R., Chaudhari, R., & Amarsinh, V. (2021). An intelligent crop price prediction using suitable machine learning algorithm. In *ITM web of conferences* (Vol. 40, p. 03040). EDP Sciences.
9. Dhanapal, R., AjanRaj, A., Balavinayagapragathish, S., & Balaji, J. (2021, May). Crop price prediction using supervised machine learning algorithms. In *Journal of Physics: Conference Series* (Vol. 1916, No. 1, p. 012042). IOP Publishing.
10. Bayona-Oré, S., Cerna, R., & Tirado Hinojoza, E. (2021). Machine learning for price prediction for agricultural products.
11. Samuel, P., Sahithi, B., Saheli, T., Ramanika, D., & Kumar, N. A. (2020). Crop price prediction system using machine learning algorithms. *Quest Journals Journal of Software Engineering and Simulation*.
12. Medar, R., Rajpurohit, V. S., & Shweta, S. (2019, March). Crop yield prediction using machine learning techniques. In *2019 IEEE 5th international conference for convergence in technology (I2CT)* (pp. 1-5). IEEE.