International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 359-363, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE58
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

Early Price Prediction of Crops Using Machine Learning Model

[1]Fathima Begum M, [2]Venkat Sai Krishna Reddy, [3]Lokesh Reddy.K

[1,2,3]Computer Science and Engineering (Cyber Security), Madanapalle Institute of Technology & Science, Annamayya District, Andhra Pradesh, India

E-mail: [1]fathimabegum.it@gmail.com, [2]bvskrishna1219@gmail.com, [3]kannapulokeshreddy82@gmail.com

***Abstract* - The agriculture sector is a cornerstone of India's economy, contributing as the third-largest GDP sector and supporting the livelihoods of millions. While farmers in India have extensive knowledge of climate conditions and crop suitability, they often face significant investment losses due to over production of specific crops, leading to low market prices. This overproduction is frequently driven by a lack of accurate price forecasting and guidance on crop selection. To address this challenge, machine learning (ML) techniques can be leveraged to predict optimal crops for cultivation and their base market prices. By analysing historical data, market trends, and environmental factors, these advanced models can provide actionable insights to farmers, help farmers increase profits, thereby reducing losses. This approach ensures informed decision-making, minimizes risks, and promotes sustainable economic growth in the agriculture sector.**

Keywords: Agriculture, Statistical, Machine learning, Prediction.

I. INTRODUCTION

India is the second most populated country in the world. Today almost 18% of world's population lives in India. Agriculture is a cornerstone of the global economy, contributing to food security, employment, and sustainable development .India's agriculture system categorized into three cropping seasons they are: 1.Rabi, 2.Kharif, 3.Zaid. These seasons are based on climate conditions and crop-growing patterns in the country. In kharif crops (monsoon crops), the growing season: June to October, harvesting September-October, Weather conditions are required lots of water and hot weather. The major crops are rice, maize, sorghum, cotton, groundnut, sugarcane, soybean, turmeric etc. Rabi crops (winter crops), the growing season: October to March, harvesting April-may, weather conditions are cool weather for growth and warm weather for ripening, Major crops are wheat, barley, peas, gram, mustard, linseed, potato, onion etc. [1]

Zaid crops (summer crops) growing season March to June and harvesting before monsoon then weather conditions are warm dry weather, short duration. Major crops are watermelon, muskmelon, cucumber, vegetables, and fodder crops. They are many factors affecting crop production mainly climate conditions, soil type, soil fertility, water quality, irrigation methods, fertilizer application, pest control, crop rotation.

Majorly the farmers are followed two types of production techniques. One is traditional methods those are like manual sowing, Traditional irrigation, organic farming. Modern methods those are like mechanized farming, drip irrigation, greenhouse cultivation, precision farming groundnuts, paddy (rice), and wheat are not classified as vegetables; they fall into different categories of crops. Groundnuts categorized to oil seeds which are rich in oil and protein. They are used for producing edible oil, and the seeds can be consumed directly as snacks or used in various culinary applications. Apart from oil production, groundnuts are used in making peanut butter, confectionery, and as an ingredient in various dishes. Paddy (Rice) categorized into cereal grain characteristics are paddy refers to the rice crop when it is still in the field or in its unprocessed form. Rice provides carbohydrates that are why it became large part in food supply, especially in Asia; it is consumed as a primary food source and is used in a variety of dishes. It is also processed into products like rice flour, rice bran oil, and rice syrup. Wheat comes under category: cereal grain and are most widely cultivated cereal grains and is a staple food in many parts of the world. It is known for its high carbohydrate content and is a significant source of energy [2].

Wheat is primarily used to produce flour, which is then used to make bread, pasta, pastries, and other baked goods. It is also used brewing and as animal feed. These crops are essential components for global agriculture and food systems, each serving distinct roles in nutrition, industry, and economy. They are not vegetables but are crucial for providing energy, protein, and essential nutrients to human diets. Actually these are foremost important to balance currency value these grains have major advantages are storage capability, government support and mechanization. Grain farming can be more easily mechanized, reducing labour costs and increasing efficiency.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 359-363, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE58
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

Now Fresh Vegetables generally have shorter growing cycles (30-90 days), allowing for multiple harvests per year quicker cash flow. Vegetables can be sold in local markets, supermarkets, or through direct-to-consumer channels like farmers' markets and CSA (Community Supported Agriculture) programs [3].

High-value vegetables like tomatoes, bell peppers, and greens chillies can yield significant profits per acre due to their market demand and price the main challenges are facing, vegetables have a short shelf life, leading losses if not sold quickly or stored properly(cold storage), market fluctuations like prices can be volatile due to seasonal oversupply or changes in consumer demand then high input costs are costs for seeds, fertilizers, pesticides, and labour can be substantial impacting profit margins. By understanding these dynamics, farmers can make informed decisions to optimize their operations, manage risks, and enhance profitability in both prices can be volatile due to seasonal over supply or changes in consumer demand then high input costs are costs for seeds, fertilizers, pesticides, and labour can be substantial impacting profit margins [4].

By understanding these dynamics, farmers can make informed decisions to optimize their operations, manage risks, and enhance profitability in both vegetable and grain farming. The fluctuations in price over a span of years is summarised in figure 1.

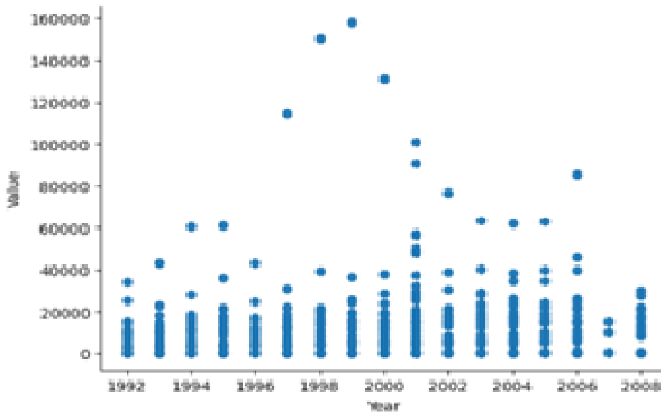

Fig. 1: Fluctuations in price over a span of years

Most of the existing research papers they are proving crop yield prediction or grain cost prediction but there are very less papers to predict vegetables price prediction but in these paper we can see mixed grains, and vegetables both prediction. Single crop improved the crop yield but targeting a single crop every year it is vulnerable to year changes pattern. Finding the crop price prediction collected data from the Food and Agriculture Organization (FAO) of the United Nations. Utilizing a comprehensive dataset, the study explores key trends in agriculture in India including production, consumption, and trade. The findings highlight significant patterns it can inform policymakers and stakeholders in the agriculture sector.

II. LITERATURE REVIEW

There are different ways to predict agricultural early price prediction through techniques like machine learning, deep learning, data analysis. They are used different approaches to predict crop price based on production. This section will discuss the procedures considering their limitations and strengths to find suitable methodologies to conduct further study. Cravero et al. [5] applied M5-prime and k-nearest neighbour (K-NN) techniques, evaluating their performance on a two-year dataset. It achieved the lowest average RMSE, and MAE alongside the highest correlation factors, demonstrating suitability for large-scale yield prediction. Despite these strengths, the dataset contained only 6,000 records, raising concerns about potential overfitting when training machine learning models. In [6] studies emphasized the integration of supervised learning techniques, such as Random Forest and Decision tree, for agricultural yield prediction, achieving robust performance with diverse environmental and soil parameters.

Challenges such as dataset limitations, feature redundancy, and overfitting persisted in 2021 prompting researchers to advocate for larger, more diverse datasets and improved validation techniques to ensure scalability and reliability. The Sorghum Yield Prediction model [7] utilizes Convolutional Neural Networks (CNN) and Linear Regression to estimate Sorghum crop yield, achieving an accuracy of 74.5% for yield prediction and an impressive 99% average precision for weight estimation. However, the dataset used was sourced from America, limiting its applicability to Indian crop yield scenarios. The rule-based crop prediction model [8], which employs association rule mining on agricultural data from 2000 to 2012, predicts yield effectively but is constrained by its focus on districts in Southern India. Notably, a common observation across studies is the superior accuracy achieved by models using the Random Forest algorithm. ARIMA models face limitations due to their inherent focus on the mean values of past series data, making them less effective at capturing rapid variations and dynamic changes in underlying processes.

In time series analysis, identifying the data series requires considering two key factors: the randomness of the data and the presence of trends. This process is followed by three critical steps: model identification, parameter estimation, and validation testing [9].

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 359-363, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE58
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

III. MATERIALS AND METHODS

A. Proposed Methodology

This section discusses the resources and techniques utilized in this work. The dataset description provides an understanding of the data and its preprocessing requirements, followed by an overview of the implemented analytical models. The proposed model predicts crop price based on a year using supervised machine learning by analysing parameters such as crop, year, production, flag description. It also suggests alternative crops based on the weather and soil conditions of the field. Among the tested models, Random Forest Regressor and Decision Tree Regressor delivered the best results for price prediction, with cross-validation scores ensuring the selection of the optimal model. For crop prediction, the Decision Tree Classifier, KNN classifier, Naive Bayes classifier, XGBoost, Gradient Boosting, and Random Forest classifier yielded the highest accuracy.

Table I: Performance Measures

Performance Metrics	Linear Regression	Random Forest
R2 score	3.12	4.19
RSD	0.44	0.36
Accuracy	84	92.3

Figure 2 illustrates the workflow of proposed method. Other supervised techniques, including Linear Regression, Multinomial Logistic Regression, and Support Vector Machines, were excluded due to their lower accuracy.

B. Dataset Attributes and Parameters: FAO Dataset

The dataset that is used to train and test early price prediction collected from Food and Agriculture Organization (FAO) is a specialized agency of the United Nations focused on achieving eradicating hunger worldwide. Established in 1945, it include 16 features like Domain Code, Domain, Area Code, Element Code, Item Code, Item, Year Code, Year, Months Code, Months, Unit, Value, Flag, Flag Description In a dataset, if rows of data having any missing values they are removed or replaced with mean value doesn't affect on the accuracy of a model. Our dataset have no missing values because it is organized data and already structured data. In a dataset, if the rows of data having any missing values they are removed or replaced with mean value doesn't affect on the accuracy of a model. Our dataset have no missing values because it is organized data and structured data. The dataset have a column called unit have three different categories. They are LCU, SLC, USD one is local market price another one is state level market price and another one is us dollar price it means international market price. Here we done convert the USD value to India currency because for reduction confusion in predicting and make accurate results [10].

C. Data Analytics and Modelling

In this section, the machine learning techniques that are separately implemented. A Predictive model for early crop price forecasting using will be explained. Crop Price Prediction Model: Predicting the continuous target or dependent variable 'yield' over the multiple input dependent features.

D. Random Forest Algorithm

Random Forest is a robust ensemble learning method that builds multiple decision trees and combines their predictions to improve accuracy and reduce overfitting. It is highly effective in dealing with categorical data, which is often complex and influenced by multiple factors such as Year, Production, price, flag description. The importance features are displayed in figure 3.

IV. EXPERIMENTAL RESULTS

The model is implemented on Python 3.7 (Goggle Collab) with the help of the following inbuilt libraries like – numpy, pandas, seaborn, scikit-learn and matplotlib. The configuration of the hardware used for the implementation is Intel i-5 or Ryzen 5 6000u minimum with 8GB RAM.

In prediction model, input data is divided into two parts one is train the model another is testing purpose the division part is 70/30 the 70 part is train the model 30 is test the model. After the implementation we compared with different machine learning algorithms because of which algorithm performance is better identification with different parameters tried and summarised in Table 1.

A. R2 Score

R-Squared Shows how much variance in a dependent variable is explained by the independent variable. The R2 Score (RSquared) measures the proportion of variance in the dependent variable that is explained by the independent variable(s).

A higher R2 score indicates that the model has effectively narrowed the gap between the actual and predicted values. The R2 score ranges from 0 to 1, with a score of 1 representing a perfect model where the independent variables fully explain the patterns in the dependent variable [11].

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 359-363, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE58
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

B. Residual Standard Deviation (RSD)

It measures the differences between the observed values and the predicted values in a regression model. It indicates the spread of residuals (errors) around the regression line.

A smaller RSD value signifies that the model's predictions are closer to the actual values, reflecting better accuracy in predicting crop Price.

V. CONCLUSION AND FUTURE WORKS

The result shows that accurate prediction multiple crop types are covered in this model making it easier to for farmers to know the price and take different decisions. In this model we mapped two different datasets and merged based on production level and then predicting the price in annual bases and state market level. And our work mainly focuses on predicting the price to make good production in less land and crop yield prediction. Future work: First, developing application for predicting crop price by collecting real world data from farmer and also from nursery's like fertilizer shops. Second, collecting day by day information for predicting price of a crop update leads to accurate results and it makes loss reduction.

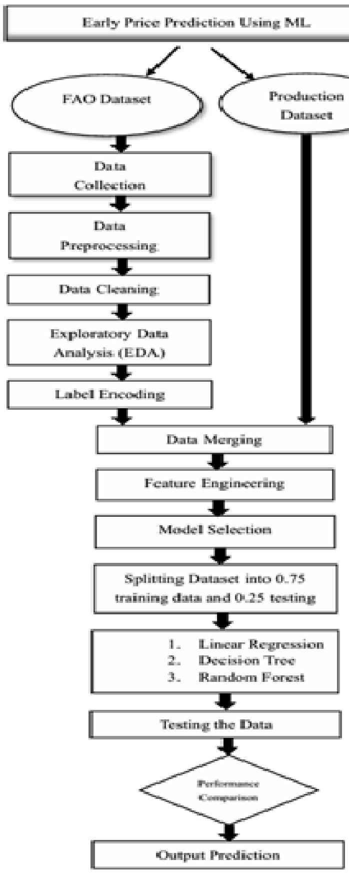

Fig. 2: Proposed workflow

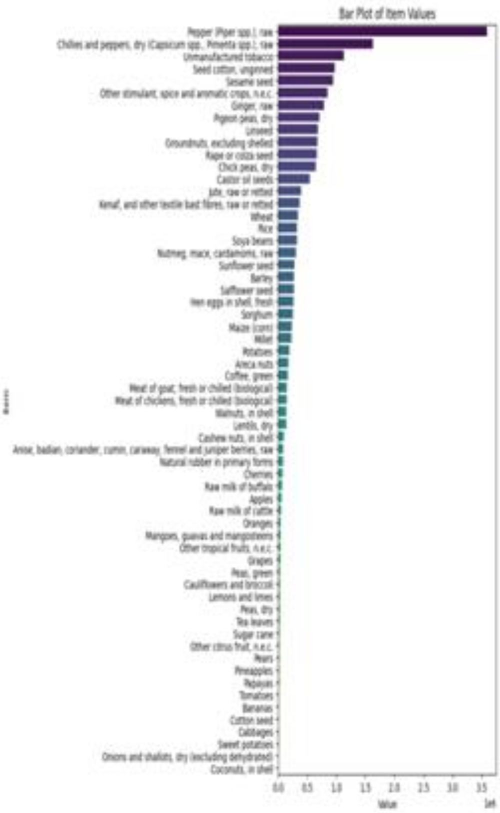

Fig. 3: Feature Importance

REFERENCES

[1] Y. Akkem, S. K. Biswas, and A. Varanasi, "Smart farming using artificial intelligence: A review," Engineering Applications of Artificial Intelligence, vol. 120, p. 105899, 2023.

[2] K. L.-M. Ang and J. K. P. Seng, "Big data and machine learning with hyperspectral information in agriculture," IEEE Access, vol. 9, pp. 36 699–36 718, 2021.

[3] N. Bali and A. Singla, "Emerging trends in machine learning to predict crop yield and study its influential factors: A survey," Archives of computational methods in engineering, vol. 29, no. 1, pp. 95–112, 2022.

[4] J. Cao, Z. Zhang, F. Tao, L. Zhang, Y. Luo, J. Zhang, J. Han, and J. Xie, "Integrating multi-source data for rice yield prediction across china using machine learning and deep learning approaches," Agricultural and Forest Meteorology, vol. 297, p. 108275, 2021.

[5] A.Cravero, S. Pardo, S. Sep´ulveda, and L. Mu˜noz, "Challenges to use machine learning in agricultural big data: a systematic literature review," Agronomy, vol. 12, no. 3, p. 748, 2022.

[6] J. Pant, R. Pant, M. K. Singh, D. P. Singh, and H. Pant, "Analysis of agricultural crop yield prediction using statistical techniques of machine learning," Materials Today: Proceedings, vol. 46, pp. 10 922–10 926, 2021.

ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 359-363, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE58
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

[7] D. Paudel, H. Boogaard, A. de Wit, S. Janssen, S. Osinga, C. Pylianidis, and I. N. Athanasiadis, "Machine learning for large-scale crop yield forecasting," Agricultural Systems, vol. 187, p. 103016, 2021.

[8] M. Rashid, B. S. Bari, Y. Yusup, M. A. Kamaruddin, and N. Khan, "A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction," IEEE access, vol. 9, pp. 63 406–63 439, 2021.

[9] D. J. Reddy and M. R. Kumar, "Crop yield prediction using machine learning algorithm," in 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2021, pp. 1466–1470.

[10] A.Sharma, A. Tamrakar, S. Dewasi, and N. S. Naik, "Early prediction of crop yield in india using machine learning," in 2022 IEEE Region 10 Symposium (TENSYMP). IEEE, 2022, pp. 1–6.

[11] H. Tong and Z. Nikoloski, "Machine learning approaches for crop improvement: Leveraging phenotypic and genotypic big data," Journal of plant physiology, vol. 257, p. 153354, 2021.

Citation of this Article:

Fathima Begum M, Venkat Sai Krishna Reddy, & Lokesh Reddy.K. (2025). Early Price Prediction of Crops Using Machine Learning Model. In proceeding of International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25), published by *IRJIET*, Volume 9, Special Issue of INSPIRE'25, pp 359-363. Article DOI https://doi.org/10.47001/IRJIET/2025.INSPIRE58

****

