



A Machine Learning Approach for Tomato Crop Yield and Price Prediction

Varsha Manohar Pujari^{*1}, Vishwanath Y^{*2}

^{1,2} School of CSE, REVA University

²vishwanath.y@reva.edu.in

*Corresponding author's E-mail: manoharvarsha96@gmail.com

Article History	Abstract
Received: 06 June 2023 Revised: 05 Sept 2023 Accepted: 29 Nov 2023	<i>Agricultural product costs play a significant part in the horticultural market. In India, vegetables, for example, tomatoes have the biggest supply and price variances among farming items. As tomatoes are grown around the year, outdoor and indoor, their yields change because of various factors, it is hard to settle tomatoes' inventory and costs. Although the Government puts numerous efforts to balance out the supply and costs of vegetables, continuous meteorological changes have prompted unstable supply and price fluctuations of vegetables. Accordingly, the right anticipating of vegetable costs is a significant issue. To oblige these, in this paper, an attempt has been made to dissect the costs and yield of tomatoes in India by utilizing a Machine Learning approach. This will unquestionably help the farmers and the Government if the anticipated costs are getting higher in the forthcoming months, then appropriate strategies can be made to diminish the costs of tomatoes.</i>
CC License CC-BY-NC-SA 4.0	Keywords: Ensemble Model, Naïve Bayes, Decision Trees, Regression, Yield Prediction, Price Prediction.

1. Introduction

Agriculture is the essential wellspring of jobs for about 58% of India's population [4]. The Indian economy is transcendently subject to the agribusiness area, which plays an important role and contributes about 17% to the absolute Gross Domestic Product, and tomato crop is one of the most significant harvests under the Indian government "TOP" need a rundown of green yields alongside Potatoes and Onions. These three TOP crops found the nation over-represents the Government of India's drive called 'Operation Greens' to improve farmers living conditions.

Tomato crop began in the Peru region of South America. India is probably the biggest producer of fresh tomatoes. It is the 2nd most significant harvest of the world after potato. In India, the tomato is cultivated in two seasons, from June- September (monsoon season) and from October to February (spring season), although in certain areas, tomatoes are developed consistently. It is utilized in, powder, puree, juice, soup, and ketchup. The significant tomato-cultivating states are Madhya Pradesh, Andhra Pradesh, Karnataka, West Bengal, Bihar, Maharashtra, Orissa, and UP. In Punjab state, Hoshiarpur, Jalandhar, Amritsar, and Ropar are tomato developing locales. It tends to be cultivated in varied soil such as sand, clay, black soil, and red soil having legitimate drainage [3].

Because of their decayable nature, tomatoes cannot be stored in their fresh state. Subsequently, tomatoes are processed to lessen their dropping rate. In India, immense post-reap losses of the harvested tomatoes happen because of deficient storage facilities, which carry generous losses to the cultivators, and henceforth it affects the country's economy. An administration statement appraises that 12.4% of tomatoes are lost [6].

The tomato preservation in a semi-processing framework deals with the attractive excess as well as guarantees the inventory of crude materials for processed items like sauce, ketchup, drink, and other handled items. Without an appropriate handling assortment, Indian processors will in general import mass tomato glue for the most part from China tomato cultivators sell their produce normally via a nearby dealer at the neighborhood or provincial market (the commercial center). Being a "TOP-crop", concerning the Government of India's main concern vegetable crops, tomatoes are dependent upon frequent fluctuations in costs & regularly ranchers face difficulties with such variances.

As it were, effective harvesting comes down to settling on difficult choices which depend on interconnections of various factors, including crop specifications, soil type, weather change, and so on. In general, farming methodologies have been applied to one of its part or a whole field. AI/ML in horticulture takes into consideration a lot higher accuracy, empowering farmers to treat animals and plants separately, which fundamentally increments the effectiveness of ranchers and the Government's decisions [5].



Fig. 1 Tomato Crop.

2. Literature Review

The authors of the paper [1], have proposed a framework to anticipate the vegetable value utilizing STL-LSTM Method. The design of the model comprises of information, preprocessing technique, and deep learning model. They considered meteorological information, and other information, and eliminated the trend and seasonality of series data by utilizing an STL preprocessing technique. An LSTM calculation was utilized to figure the month-to-month costs of vegetables.

In [2], the authors have proposed a model which predicts the costs of fruits and vegetables of forthcoming months in Bangalore. They have applied Seasonal ARIMA to anticipate the costs of fruits and vegetables in the coming months. The study gives whether the costs of natural products i.e., Mosambi, Mango, and pineapple and vegetables i.e., Brinjal, Ladyfinger, and cauliflower have expanded over years or not and furthermore important estimates that can be taken all together to make products moderate.

In paper [7], They developed a crop recommendation model that uses Artificial Neural Network, which can be used by farmers all over India. This model would help the farmers in taking right decision about which crop to grow depending on various meteorological factors. This model also has another system, called Rainfall Predictor, which predicts the rainfall for coming months.

In paper [8], they have examined cost of vegetable approach introduction regarding the cost expectation in farming, a different overview about data extraction strategies in the farming land for price forecast, and time series analysis of the estimated cost. Three fundamental day-by-day usage vegetables are taken for the study. The primary objective of this framework is to give the client and farmers the exact and in-time commodity cost at the right time.

In paper [9], they have proposed a model in which the future demand of the harvests are anticipated. The total crops cultivated, and the general utilization of each crop are identified from the dataset dependent on attributes, for example, crop territories, sorts of yield cultivated, nature of the soil, yields and the general harvests devoured gathered from the existing dataset. The recommended information may contain one or more crops depending on the higher demand grade.

proposed work

This work focuses on the prediction of tomato crop prices, tomato crop yield, and whether the crop can be cultivated or not.



Fig. 2 Tomato Farmland

3. Materials And Methods

This section gives information about the dataset and methodologies used to predict the crop prices, crop yield, and whether the crop can be cultivated or not.

A. Data Exploration

- 1) *Acquisition of Dataset:* Data for prediction of tomatoes such as prices, month, and location were taken from agmarknet.in [19]. The data we obtained was from 2015 to February 2021. Dataset also consists of various attributes such as Type of soil, Temperature, Humidity, season, area, crop yield, and whether the crop can be cultivated or not.

B. Data Cleaning and Data preparation

The presence of null and empty values in the dataset decreases the value of the information and also accordingly influences the functioning of AI system. Hence, in order to deal with these empty values, the null values present in the dataset are dropped and categorical data such as Type of soil, Season, Month, and Location have been encoded from 0,1,2,3,.....,n by using a label encoder to generate class labels [10-25].

C. Data Modelling

- 1) *Training and Testing:* We have considered 80% training data and 20% testing data.
- 2) *Algorithms:* In this model algorithms such as Naïve Bayes, Decision Tree, Ensemble model, and Linear Regression for the prediction of tomato prices, tomato crop yield and to predict whether the crop can be cultivated or not.
- 3) *Naïve Bayes Algorithm:* Naïve Bayes is a simple algorithm that is used for classification. This algorithm works on the concept of Probability i.e., to predict tomato can be cultivated or not (YES or NO).
- 4) *Linear Regression:* Regression is an algorithm in which the output to be predicted is dependent on the given input variable. In this work, regression is used to predict the tomato yield and the prices.
- 5) *Ensemble Method:* The Ensemble method uses multiple models to give better accurate results. In this work, Naïve Bayes and Decision Tree are the models used in Ensemble method for classifying whether the crop can be cultivated or not i.e., YES or NO.

D. Data Evaluation

- 1) *Trained ML Model and Predictions:* After feeding the data to ML algorithms, we obtained the predictions for crop yield and prices.

We have also used Tkinter through which the user can enter the data for the predictions. Tkinter is a GUI module for Python. When Python is integrated with the Tkinter module, it provides an easy and efficient way to create GUI applications.

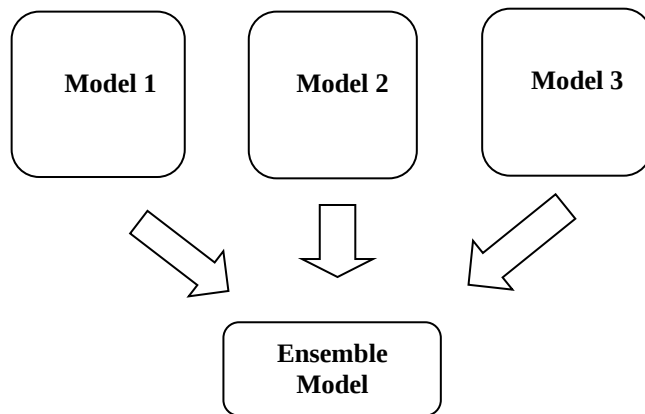


Fig. 3 Ensemble Concept

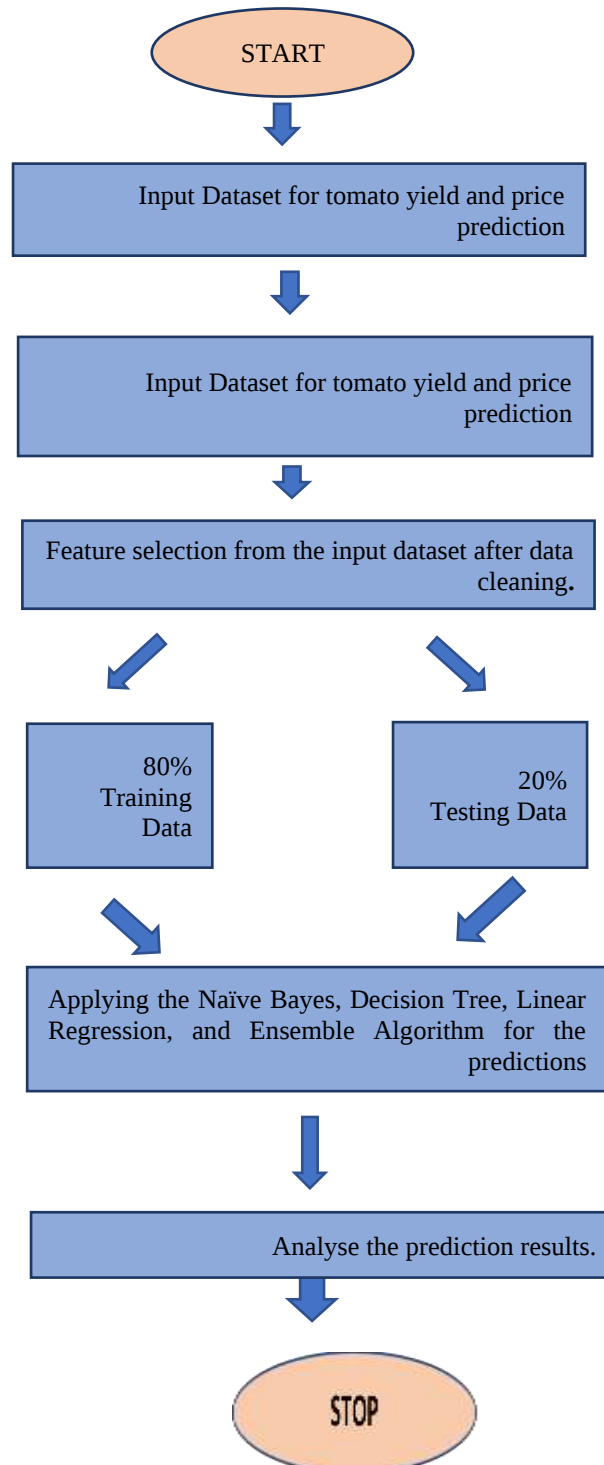


Fig. 4 Flow Chart

3. Results and Discussion



The image shows a Tkinter window titled "Tomato Crop and price prediction". It contains a grid of input fields for "TypeofSoil", "Temperature", "Humidity", "Month", "Season", "Location", and "area". To the right of these fields are output labels: "Crop", "yield", and "price". At the bottom, there are three buttons labeled "predict Crop", "predict yield", and "predict price".

Fig. 5 Tkinter UI Window

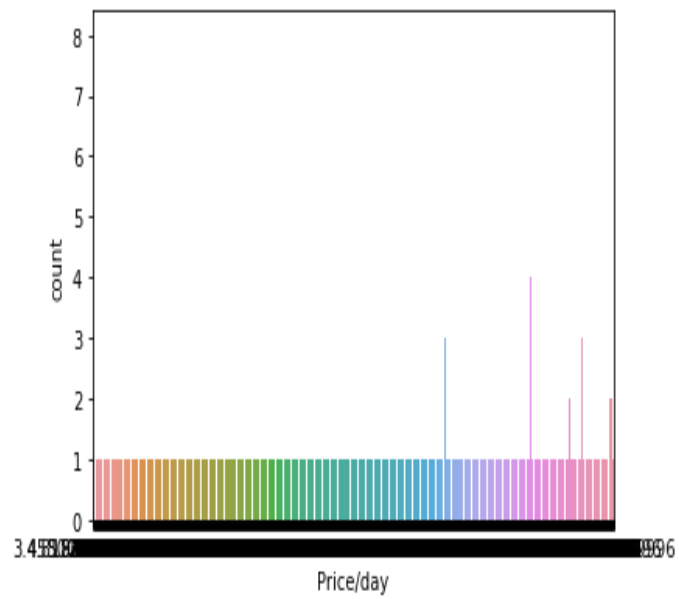


Fig. 6 Tomato Price Graph

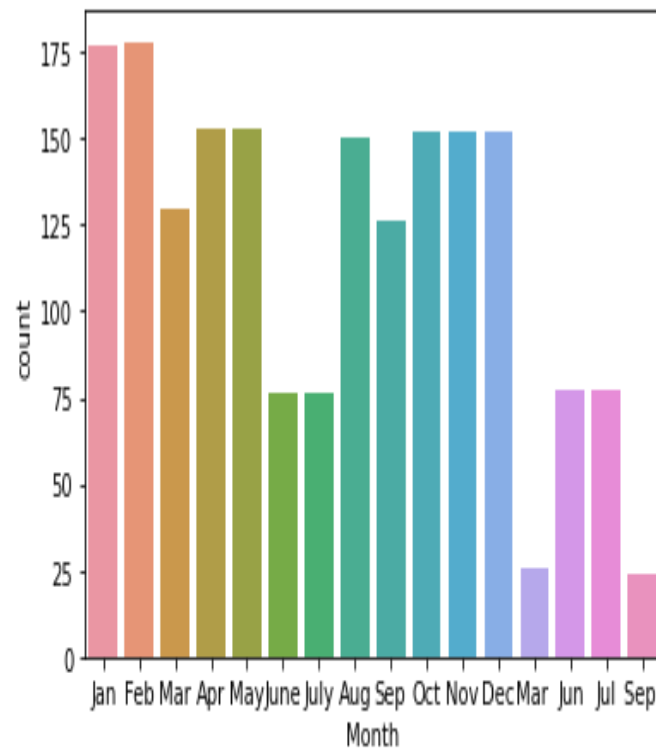


Fig. 7 Month-wise Crop Yield

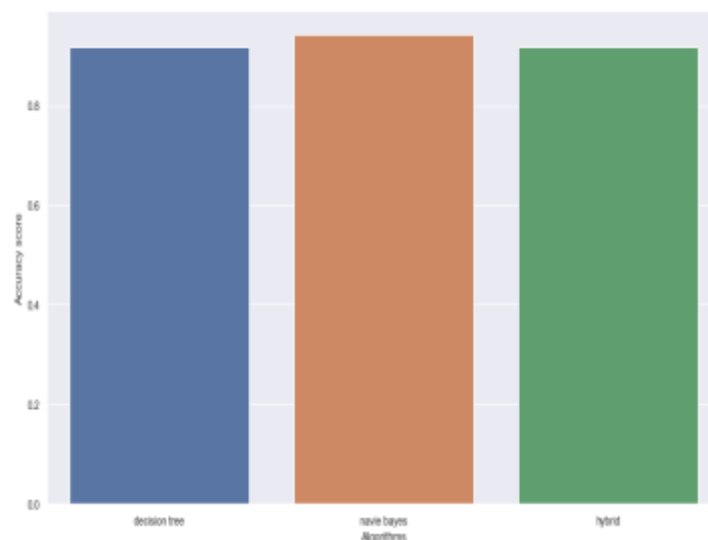


Fig. 8 Algorithm accuracy for classification

4. Conclusion

This work mainly focuses on the tomato crop yield, price and whether the tomato can be cultivated or not. This model works well with small dataset. Dataset consists of data and tomato prices from the year 2015- Feb 2021. Graphs depicts, in which month the prices, tomato yield is high, and the accuracy of algorithms for classification. This system would help the Indian farmers and the Indian government to take proper and make effective decisions.

The work proposed in this paper can be further implemented in the future to predict fertilizer suitable for tomato crop yield. The crop yield and price prediction can also be done by collecting the agricultural farm- land images clicked using a camera or a drone. This can also be integrated with an IOT device to make it more efficient and user friendly.

Acknowledgement

I express gratitude toward **Dr. Vishwanath Y** for his skill and help all through all parts of our study and work on "A Machine Learning Approach for Tomato Crop Yield and Price Prediction" and for his assistance in writing the paper.

References:

1. Dong Jin, Helin Yin, et al. "Forecasting of Vegetable Prices using STL-LSTM Method." The 2019 6th International Conference on Systems and Informatics (ICSAI 2019), 978-1-7281-5255-1 2019 IEEE.
2. Dharavath Ramesh, Ekaansh Khosla. "Seasonal ARIMA to forecast fruits and vegetable agricultural prices.", 2019 IEEE International Symposium on Smart Electronic Systems.
3. Tomato Farming, Apni Kheti website. [Online]. Available: <https://www.apnikheti.com/en/pn/agriculture/horticulture/vegetable-crops/tomato>
4. Agriculture in India: Information About Indian Agriculture & Its Importance, India Brand Equity Foundation website. [Online]. Available: <https://www.ibef.org/industry/agriculture-india.aspx>
5. Machine Learning in Agriculture: What It Can Do Now and in the Future, Yaroslav Kufilinski, Iflexion website. [Online]. Available: <https://www.iflexion.com/blog/machine-learning-agriculture>
6. India: The sleeping tiger? - Part 1, Sophie Colvine, Tomato News website. [Online]. Available: http://www.tomatonews.com/en/india-the-sleeping-tiger--part-1_2_1213.html
7. Zeel Doshi, Subhash Nadkarni, et al. "Intelligent Crop Recommendation System Using Machine Learning Algorithms". 978-1-5386-5257-2, 2018 IEEE.
8. R. Deepalakshmi, S. Padma Devi, et al. "An algorithm for dynamic price prediction of vegetable commodities using statistical price prediction technique". Int. Journal. Computational Complexity and Intelligent Algorithms, 2019.
9. Kanaga Suba Raja, Rishi R, et al. "Demand based crop recommender system for farmers". 2017 IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development.
10. A. Amarendra Reddy, "Price Forecasting of Tomatoes". International Journal Of Vegetable Science 2018.
11. Gan-Qiong Li, Shi-Wei Xu, and Zhe-min Li, "Short-term Price Forecasting for Agro-items Using Artificial Neural Networks", International Conference on Agricultural Risk and Food Security 2010.
12. Nasira and Hemageetha, "Forecasting Model for Vegetable Price Using Back Propagation Neural Network", International Journal of Computational Intelligence and Informatics, September 2012.
13. Girish K. Jha and Kanchan Sinha, "Agricultural Price Forecasting Using Neural Network Model: An Innovative Information Delivery System", Agricultural Economics Research Review, December 2013.
14. Jianhua Zhanga, Fantao Kongb, Jianzhai Wu, Mengshuai Zhu, Ke Xu, and JiaJia Liu, "Tomato prices time series prediction model based on wavelet neural network", Trans Tech Publicationa, 2014.
15. Rajeswari and Suthendran, "Developing an Agricultural Product Price Prediction Model using HADT Algorithm", IJEAT, December 2019.
16. Hemageetha and Nasira, "Vegetable Price Prediction using Adaptive Neuro-Fuzzy Inference System" International Journal of Computer Sciences and Engineering, 2017.
17. Sagar Pathane, Uttam Patil, Nandini Sidnal, "Prediction of Future Market Price for Agricultural Commodities". International Journal of System and Software Engineering 2015.
18. Balaji Prabhu B V and Dr. M Dakshayini, "Demand-Prediction Model for Forecasting AGRI-Needs of the Society". Proceedings of International Conference on Inventive Computing and Informatics (ICICI 2017).
19. Agmarknet, Government of India website. [Online]. Available: https://agmarknet.gov.in/PriceTrends/SA_Pri_Month.aspx
20. Sudhan Murugan Bhagavathi, Anitha Thavasimuthu, Aruna Murugesan, Charlyn Pushpa Latha George Rajendran, A Vijay, Raja Laxmi, Rajendran Thavasimuthu, Weather forecasting and prediction using hybrid C5.0 machine learning algorithm International Journal of Communication Systems, Vol. 34, Issue. 10, Pp. e4805, 2021.
21. PM Surendra, S Manimurugan, A New Modified Recurrent Extreme Learning with PSO Machine Based on Feature Fusion with CNN Deep Features for Breast Cancer Detection, Journal of Computational Science and Intelligent Technologies, Vol. 1, Issue. 3, Pp. 15-21, 2020.
22. PK Sadineni, Comparative Study on Query Processing and Indexing Techniques in Big Data, 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp. 933-939, 2020.
23. AH Omar Baabood, Prajoona Valsalan, Tariq Ahmed Barham Baomar, IoT Based Health Monitoring System, Journal of Critical Reviews, Vol. 7, Issue. 4, pp. 739-743, 2020.
24. Sajay KR, Suvanam Sasidhar Babu, Vijayalakshmi Yellepeddi, Enhancing The Security Of Cloud Data Using Hybrid Encryption Algorithm, Journal of Ambient Intelligence and Humanized Computing, 2019. <https://doi.org/10.1007/s12652-019-01403-1>
25. Bindhia K Francis, Suvanam Sasidhar Babu, Predicting academic performance of students using a hybrid data mining approach, Journal of Medical Systems, 43:162, 2019. <https://doi.org/10.1007/s10916-019-1295-4>